BEHAVIORAL AND BRAIN SCIENCES (1998) **21**, 1–54
Printed in the United States of America

The development of features in object concepts

Philippe G. Schyns
Department of Psychology, University of Glasgow, Glasgow, G12 8QB United Kingdom
Electronic mail: *philippe@psy.gla.ac.uk* *www.gla.ac.uk/acad/psychology*

Robert L. Goldstone
Department of Psychology, Indiana University, Bloomington, IN 47405
Electronic mail: *rgoldsto@ucs.indiana.edu* *cognitrn.psych.indiana.edu/*

Jean-Pierre Thibaut
Department of Psychology, Université de Liège, Batiment B32, Sart-Tilman 4000 Liège, Belgium
Electronic mail: *jthibaut@ulg.ac.be*

Abstract: According to one productive and influential approach to cognition, categorization, object recognition, and higher level cognitive processes operate on a set of fixed features, which are the output of lower level perceptual processes. In many situations, however, it is the higher level cognitive process being executed that influences the lower level features that are created. Rather than viewing the repertoire of features as being fixed by low-level processes, we present a theory in which people *create* features to subserve the representation and categorization of objects. Two types of category learning should be distinguished. Fixed space category learning occurs when new categorizations are representable with the available feature set. Flexible space category learning occurs when new categorizations cannot be represented with the features available. Whether fixed or flexible, learning depends on the featural contrasts and similarities between the new category to be represented and the individual's existing concepts. Fixed feature approaches face one of two problems with tasks that call for new features: If the fixed features are fairly high level and directly useful for categorization, then they will not be flexible enough to represent all objects that might be relevant for a new task. If the fixed features are small, subsymbolic fragments (such as pixels), then regularities at the level of the functional features required to accomplish categorizations will not be captured by these primitives. We present evidence of flexible perceptual changes arising from category learning and theoretical arguments for the importance of this flexibility. We describe conditions that promote feature creation and argue against interpreting them in terms of fixed features. Finally, we discuss the implications of functional features for object categorization, conceptual development, chunking, constructive induction, and formal models of dimensionality reduction.

Keywords: concept learning; conceptual development; features; perceptual learning; stimulus encoding

1. Introduction

We believe that an influential and powerful idea in cognitive science must be revised in order to provide a full account of cognition. This idea is that cognitive processes such as categorization and object recognition operate on a fixed set of perceptual or conceptual features, which are the building blocks for complex object representations. We will argue that categorization and object recognition often require the creation of new features. The featural repertoire, rather than being fixed, is dependent on situational demands, novel categorization requirements, and environmental contingencies.

In this target article, a *feature* will refer to any elementary property of a distal stimulus that is an element of cognition, an atom of psychological processing. This does not imply that people are consciously aware of these properties. Instead, features are identified by their functional role in cognition; for example, they allow new categorizations and perceptions to occur. Stimulus dimensions are ordered sets of feature values, such as size, brightness, and hue. Two features can create a new stimulus dimension, for example, by interpolating the intermediate values between poles defined by the two features.

1.1. Fixed feature vocabularies

In a typical application of the fixed features approach to categorization (see, e.g., Bruner et al. 1956), subjects are shown simple objects and are instructed to learn the rule for sorting them. Such rules are based on logical combinations of features that are manifestly present in the stimuli. For example, a subject might learn a rule that objects that are *white* and *square* should be put in the same category. The subject does not have to create the relevant features to be used for categorization. Instead, there is an implicit agreement between the experimenter and the subject about what features compose the stimuli.

Although categorization research has come a long way since these early experiments, many recent approaches to categorization have continued to use stimuli that "wear their features on their sleeves." Clear-cut dimensions with distinct values are often used for reasons of experimental hygiene. Researchers have used simple shapes (Murphy & Ross 1994), line positions (Aha & Goldstone 1992), colors (Bruner et al. 1956), and line orientations (Nosofsky 1987) as the sources of variation in their experiments. This tendency to compose stimuli out of components has also influenced other fields. In the recognition by components (RBC) theory (Biederman 1987), combinations of a fixed set of 36 geometric elements are used to account for the recognition of a very large set of objects. Theories of phoneme (Jacobson et al. 1963) and letter (Gibson 1971; Selfridge 1959) recognition are also based on a limited set of primitives. Schank's (1972) conceptual dependency theory likewise postulates a fixed set of about 20 semantic primitives such as PTRANS (physical transfer) and INGEST (cf. Katz & Fodor, 1962, for a related point of view). This work varies widely in the kinds of components used and how they are combined, but all theories assume that representations are composed from a fixed feature set.

PHILIPPE G. SCHYNS lectures psychology at the University of Glasgow. He researches the interactions between the organization of categories in memory and the perception of the input. His current research focuses on the idea that the abstract information requirements of a categorization task can change the perception of complex visual stimuli, including faces, objects and scenes.

ROBERT GOLDSTONE is an Associate Professor of Psychology and Gill Fellow of Cognitive Science at Indiana University. Professor Goldstone's research on concept learning, perception, computational models of cognition, comparison processes, and reasoning was twice awarded American Psychological Association Division of Experimental Psychology Young Investigator Awards in 1995. He received the 1996 Chase Memorial Award for outstanding young researcher in cognitive science, and the 1991 Marquis award for outstanding psychology dissertation from the University of Michigan. He is on the editorial board of *Journal of Experimental Psychology: Learning, Memory, and Cognition.* His current research focuses on interactions between perceptual and conceptual learning.

JEAN-PIERRE THIBAUT works at the University of Liege where he lectures cognitive development. He got his PhD in 1991 and holds a degree in speech pathology. His research focuses on conceptual development. Specifically, he studies the abstraction of relevant features and their generalization. His research also deals with the encoding of implicit information during concept learning.

Fixed features are the lowest level building blocks of object representation and categorization. Any functionally important difference between objects must be representable as differences in their building blocks if it is to be represented within the system. It is typically assumed that these features are nondecomposable units or "atoms," although, if pressed, many researchers would concede that their atoms may be decomposable if necessary. All the strengths of "mental chemistry" are inherited by this approach: a very large number of object descriptions can be generated from a finite set of elements and a set of combination rules. In addition, combinations of features allow for structured hierarchical representations (Palmer 1977), as opposed to the template approach to recognition (Ullman 1989), which typically does not assume a decomposition into building blocks. In addition, the systematic relations between different objects can be expressed in terms of their features and their combination rules (Fodor & Pylyshyn 1988).

We wish to retain these powerful properties of componential representations, but we also wish to provide a framework for augmenting feature sets with new features. Componential theories of cognition should provide ways to develop new representations. Fixed feature theories limit new representations to new combinations of the fixed features. Consequently, all possible categorizations are bounded by the possible combinations of the features. If a categorization requires a feature not present in, or derivable from, the feature set, then the categorization cannot be learned. This is a rather limiting view of representational change. There may be occasions when features not originally present in the system are useful for distinguishing between important categories in the world that newly confront the organism. A system that is constructed flexibly enough to learn such features would be able to tailor its feature repertoire to the demands of categorization. In many situations, it is unrealistic to think that a system comes fully equipped to deal with all possible contingencies in a complex environment.

We will provide an account of feature learning in which the components of a representation have close ties to the categorization history of the organism. We will discuss the empirical evidence suggesting that such development occurs, and the reasons why learned features are necessary. Although we will not propose a specific implementation of flexible feature learning, we will discuss computational models that can account for learned features and how current models must be supplemented. Our analysis is addressed to literatures on both object recognition and categorization. Although these fields have not traditionally been linked, both deal with the question "What is this object?" To recognize an object as a cart is equivalent to placing the object into the category of things called "carts." In both cases, the problem is to detect the relevant features of the object in the visual array.

1.2. Empirical evidence for learned features

Although it is not addressed by fixed feature approaches to categorization, there is evidence that substantial changes occur to perceptual systems during learning. The most parsimonious explanation for some of these perceptual changes is that structures in the environment are discovered and incorporated as new features of psychological processing. This is what we mean by "feature creation."

Before we review the evidence, a few ground rules are needed. First, we distinguish between feature weighting and feature creation. A feature that is useful (diagnostic) for a categorization may be selectively attended. This selective attention may be simply a decisional strategy that does not affect the appearance of the object to be categorized (Elio & Anderson 1981; Nosofsky 1987). For example, to categorize efficiently, Elio and Anderson's subjects learned to base their judgments on diagnostic features even though they could still easily perceive the nondiagnostic features. Some researchers, however, have hypothesized that features are selectively weighted if they are diagnostic, and that this selective weighting affects perceptual, rather than only strategic or final decisional processes (Gibson 1969). In both of these views, it is assumed that the changes are based on previously existing features or dimensions. A third view is possible, in which new features or dimensions are created in the service of categorization requirements. The creation of new features is implied if the required number of prespecified features would otherwise be implausibly large.

Second, the reported experiments that we review differ in the level at which representational change is assumed to take place. Sometimes, representational changes are relatively late and strategic. Learning may consist of coming to use a previously diagnostic feature under new conditions (Lawrence 1949). On other occasions, feature changes are relatively perceptual and nonstrategic. Relevance for categorization may influence relatively perception-based tasks. It is notoriously difficult to draw a sharp distinction between perceptual and conceptual tasks; we will argue that it is ill-advised to make the distinction. For example, same–different judgment tasks (tasks in which subjects are required to judge whether or not two simultaneously presented stimuli are physically identical) have usually been thought of as providing relatively clear evidence for perceptual similarity. However, to the extent that subjects always have to represent, remember (albeit for a very short time), and attend to aspects of the compared stimuli, we cannot be certain that these tasks tap purely sensory representations. Still, by examining the particular stimuli and task demands, we might be able to assess the relative contributions of strategic and perceptual factors.

1.2.1. Preexposure. The simplest form of perceptual learning that has been studied is predifferentiation (Gibson & Walk 1956). In predifferentiation, exposure to stimuli before testing results in heightened sensitivity to those stimuli. For example, human subjects are better able to distinguish between "doodles" (a contiguous concatenation of randomly selected complex curves) after repeated exposures to them. Researchers (see, e.g., Gibson 1991) have interpreted preexposure results as perceptual differentiation, a process in which aspects of the stimuli that serve to distinguish them are made more salient. Feedback on the classification or use of stimuli is not required for sensitization; simple exposure to the stimuli suffices.

1.2.2. Diagnosticity-driven learning. Although preexposure effects indicate that category feedback is not a prerequisite for learning new aspects of the stimuli, other studies have suggested that categorization exerts an additional influence on how subjects deal with the stimuli. Subjects become selectively attuned to diagnostic features that facilitate discrimination between categories. Lawrence (1949) described a theory of acquired distinctiveness in which the cues relevant to a task become more differentiated. For example, rats were rewarded for choosing one stimulus over another in a rough–smooth discrimination task. Subsequently, the rats were tested on a discrimination task in which, for example, rough patterns required left responses and smooth patterns required right responses. Rats learned this second discrimination more quickly than rats who were first given a black–white discrimination.

Although experiments of this sort show that dimensions can be selectively sensitized, they provide little evidence for perceptual changes per se. One simple explanation of these results is that the organism simply generalizes the usefulness of a dimension from one situation to another. However, other recent data suggest that categorization diagnosticity (the predictability of a category from its building blocks) influences an object's representation in terms of features. It can affect perceptual changes in at least two ways. First, it can influence the discriminability of values within existing dimensions, or the discriminability of entire preexisting dimensions. For example, Goldstone (1994a) gave human subjects categorization training on squares varying in size or brightness. After prolonged training, subjects were tested in a same–different task. When a dimension had been relevant for categorization, same–different judgments along this dimension were more accurate (using the d′ measure from signal-detection theory) than those of subjects for whom the dimension had been irrelevant or those of control subjects who had not undergone categorization training. The greatest acuity increase along the categorization-relevant dimension was found between those points that had served as the boundaries between the learned categories. However, this sensitization of the relevant dimension also extended to other points along that dimension even though those were originally placed in the same category. In addition, one case of *acquired equivalence* was found in which discrimination along a dimension that was *irrelevant* for categorization became less acute than it was in control subjects. Because same–different judgments involve "cognitive" factors such as (very) short-term memory, attention, and encoding, these results do not guarantee that the changes were *perceptual,* but at least it can be said that the categorization training influences a task that many researchers have assumed to tap relatively low-level perceptual processes. Andrews et al. (1997) have found similar influences of categorization on similarity judgments.

Category diagnosticity can also influence perception by participating in the creation of new features for object categorization. For example, Schyns and Murphy (1991; 1994) provide evidence for such a process. In a typical experiment, subjects had to learn to label new objects and were later tested on the features used to encode the categories. The stimuli were continuous, three-dimensional, rock-like "blobs" (see Fig. 3a). The stimuli had a complex blob structure so that naive subjects showed little agreement in how they decomposed them before categorization training. The categories were defined by a coherent group of a few contiguous parts present in each category member; the rest of the blobs of an object had random shapes. After learning to categorize them, subjects were instructed to decompose the objects into parts that they thought were relevant. These parts tended to be the ones that were diagnostic for categorization. This parsing differed from what it had been before the training and occurred despite a strong bottom-up constraint (the minima rule; Hoffman & Richards 1984) on object parsing that would predict parsings other than those obtained in the experiments. Schyns and Murphy's subjects did their parsing by outlining the parts of each object (using either a computer mouse or a pen). Although it is not free of cognitive influences, this technique has the advantage of leaving subjects free to report any fragment of a stimulus they wish (independently of whether it has an easily expressible name). Braunstein et al. (1989) found that an outlining method gave the strongest evidence for parsing with the minima rule, so Schyns and Murphy should have found evidence of physically determined parsing with this task; instead, parsing was determined mostly by categorization constraints.

For the Martian rocks experiments, hypothesizing that the effects are due to shifts of attention to existing features would require positing an implausibly large number of dimensions or features. Explanations in terms of mechanisms for dynamically creating new features seem more parsimonious in these cases.

1.2.3. Differentiation. Several researchers have suggested that experience with stimuli results in subjects differentiating stimulus dimensions that were originally processed together. There is substantial developmental evidence that children are more likely to perceive stimuli in an undifferentiated manner, whereas adults analyze the stimuli into distinct dimensions. For adults, some pairs of dimensions, such as size and brightness, are called "separable" (Garner 1974). They are processed separately; attention can be selectively placed on just one of the dimensions, and similarities between stimuli are computed by summing their separately determined dimensional differences. Other pairs of dimensions, such as the saturation and brightness of a color, are called "integral." Such dimensions appear to be psychologically "fused," in that it is difficult to attend selectively to just one of them, and similarities between stimuli are computed by considering the two dimensional differences simultaneously. Several studies have indicated that children process separable dimensions in the same way in which adults process integral dimensions (Smith & Kemler 1978; Ward 1983). One explanation for these results is that part of the maturation process is to separate dimensions that were not originally separated. Such a process has also been implicated for learning distinctions between more abstract dimensions such as *heat* and *temperature* (Smith et al. 1985). The differentiation of dimensions seems to occur even in adulthood. Through training, the saturation of a color can be psychologically differentiated from its brightness (Burns & Shepp 1988; Goldstone 1994a).

There is a second type of differentiation in which categories, rather than dimensions, are split apart. Developmental studies find that the lexical categories of young children are often broader than the lexical categories of adults (Clark 1973). For example, when children overgeneralize category labels, they may group together all round objects as instances of "ball" (Chapman et al. 1986). Eventually, after a progressive reorganization of their concepts, children's lexical categories narrow and match those of adults. Adding features to an initially broad concept presumably allows it to be differentiated into more specific concepts. The acquisition of new features more specifically tuned to the categorization tasks at hand may also underlie the development of adults' conceptual expertise. Tanaka and Taylor (1991) studied categorizations by dog and bird experts. Experts were particularly adept at making fine discriminations within their category of expertise, suggesting that they acquire features specific to their domain of expertise. Schyns (1991) provided a neural network model of this type of conceptual differentiation. In a two-layered net, units initially representing a broad category became progressively specialized in representing finer categories on the basis of a feature-extraction process.

1.2.4. Summary. The experimental evidence reviewed above indicates that our categorizations, rather than being based on existing perceptual features, determine the features that enter the representation of objects. Some perceptual changes may arise from mere exposure to objects, but others depend on the way in which objects of the environment are organized into categories. In addition, perceptual dimensions and categories both undergo a differentiation process based on environmental contingencies. These results provide an initial indication that categorical constraints could influence features: Rather than being fixed and unaffected by experience, features could be progressively extracted and developed as an organism categorizes its world.

2. A functional approach to feature creation

2.1. The function of features

The function of a feature is to mark commonalities between members of the same category and to distinguish between categories. In fixed feature approaches to categorization and object recognition, a functional constraint guides the construction of the repertoire of features. Researchers develop their feature sets by keeping in mind the question "What features would be required to solve this categorization task?" In many cases, the researchers then test their theories using stimuli that were constructed from these feature sets.

We agree that features should be functionally determined. However, their constraints should be defined by the environment and not simply by the experimenter. Even if the fixed feature researcher manages to draw upon a plausible feature set, it will probably be limited to a specific domain and will not adapt to temporary or local environmental states. More importantly, to restrict research to the problem of combining obvious, clearly demarcated features is to oversimplify the task of categorization.

Consider the current object recognition literature, in which Biederman's geon theory of object recognition (Biederman 1987) is contrasted with the multiple-views approach (Edelman & Bülthoff 1992; Poggio & Edelman 1990; Tarr & Pinker 1989). In recognition by components (Biederman 1987), objects are represented by a set of geometric elements derived by taking various geometric slices through the possible transformations of a generalized cone. The resulting elements can be distinguished from each other on the basis of a few nonaccidental features, features that are invariant over a wide range of transformations (rotational, translational, and scalar). Transformational invariance is a desirable property; telephones do not change their category membership (the fact that they are telephones) simply because they are rotated. However, Biederman's feature set is severely limited in its application to many natural objects (Kurbat 1995; Ullman 1989); it does not allow discriminations between many similar categories, and objects within the same category will not necessarily be represented by the same geon structure. These limitations are problems not for Biederman's theory alone but also for any approach that cannot adapt its building blocks flexibly to categorical constraints.

For example, recent object recognition research has demonstrated that the relationship between the observer and the object influences recognition performance (see, e.g., Edelman & Bulthoff 1992; Palmer et al. 1981; Tarr & Pinker 1989). Viewpoint-dependent recognition was interpreted by Tarr and Pinker (1989) as evidence that objects are represented in memory by a collection of specific views (see also Poggio & Edelman 1990). When views of an object are the basis of object representation, it is difficult to determine which subset of the set of all possible views best predicts categorizations. Categorizations are so diverse that there may not be a unique, canonical, task-independent, view-based representation of a particular object (Hill et al. 1997; Schyns, in press). For example, any view of your face could reveal diagnostic information to distinguish it from a car, but fewer views would be well suited to discriminate your face from a face of the other sex, and still fewer views would reliably distinguish your face from another face of the same gender. Viewpoint dependence appears to be relative to the diagnostic information in the task considered and to the location of the information on the object.

Hence, both the geon- and the view-based approaches to object recognition must tune their representations to the functional roles of their building blocks. That is, both theories must consider the possible categorizations of an object before considering the possible geometric elements or views that will be used to represent the object.

2.2. Categorical constraints on feature creation

A categorical context is composed of the categories and features individuals know at a particular point in their conceptual development, plus the new category to be encoded. The individual knows what the categories are from external feedback, that is, the consequences of their miscategorizations. Contrary to the classical assumption that category learning operates on fixed features, we suggest that features are flexible; that is, they adjust to the perceptual experience and the categorization history of the individual. Flexible features open the possibility that the same input is differently perceived and analyzed before being categorized. Hence, a complete theory of conceptual development should not only explain the ways in which object features are combined to form concepts, it should also explain the development of the features participating in the analysis of the input.

The role of categorization constraints on feature creation was explicitly studied by Schyns and Rodet (1997) using three categories of unknown stimuli called "Martian cells" (see Fig. 1). Categories were defined by specific blobs common to all category members to which irrelevant blobs were added (to simulate various cell bodies). Figure 1 shows, from left to right, an exemplar of the *XY* category, an exemplar of *X*, and one of *Y*. The figure also shows the features *x*, *y*, and *xy* defining each category. Note that *xy* is the conjunction of *x* and *y*.

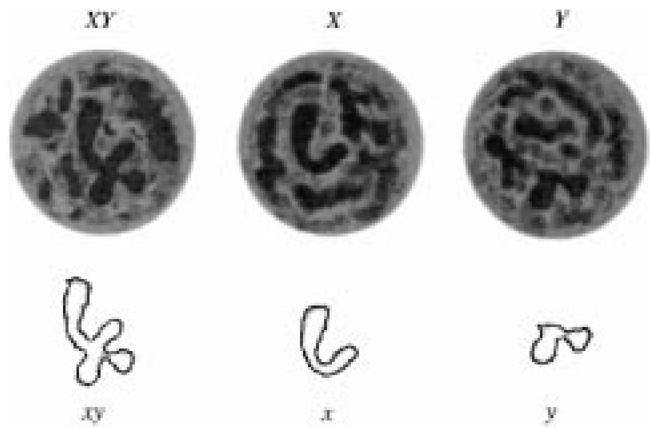

Figure 1. Design of Schyns and Rodet's (1997) feature creation experiment. From left to right, the top pictures are Martian cell exemplars from the *XY*, *X*, and *Y* categories. From left to right, the bottom pictures are the features *xy*, *x*, and *y*, defining the categories. Note that the feature *xy* is a combination of feature *x* and feature *y*. Subjects in the $XY \rightarrow X \rightarrow Y$ (vs. $X \rightarrow Y \rightarrow XY$) group learned the category in this order.

The main goal of Schyns and Rodet's experiment was to demonstrate that different categorization constraints could induce orthogonal perceptions of the defining component of *XY*, that is, perceptions of *xy* as an *x* plus *y* feature conjunction or as an *xy* unitary feature. One group of subjects was asked to learn *X* before *Y* before *XY*($X \rightarrow Y \rightarrow XY$); the other group learned the same categories in a different order ($XY \rightarrow X \rightarrow Y$). Reliable classifications of *X*, *Y*, and *XY* stimuli in the testing phase indicated, without any doubt, that all subjects saw and attended to the components *x* and *y*. *X-Y* cells were used to understand the perceptual analysis of *XY*. *X-Y* cells were *XY* exemplars in which the *x* and *y* components were not adjacent to each other. The reasoning was that subjects should categorize *X-Y* cells as *XY* members if they perceived and represented *XY* as a conjunction of two individuated features. Results revealed that only one group ($X \rightarrow Y \rightarrow XY$) performed this categorization; the perception of *XY* in the other group prompted *X* or *Y* classifications of *X-Y*. In sum, orthogonal classifications of *X-Y*, when its component features were both clearly perceived and used in the experimental groups, suggested that different features were acquired to analyze perceptually and to represent *XY*. Network simulations further suggested that the feature vocabularies were $F_{X \rightarrow Y \rightarrow XY} = \{x,y\}$ and $F_{XY \rightarrow X \rightarrow Y} = \{xy,x,y\}$, respectively. Together, these results challenge the main claim of fixed feature approaches that category learning consists only of weighting the features of a fixed set that tend to characterize categories. It appears that category learning also changes the features used perceptually to analyze the input.

Rodet and Schyns (1994) also tested more specifically the role of the context of categorization on the perceived similarity of stimuli. In the first part of Rodet and Schyns's Experiment 3, two groups of subjects learned two Martian cells categories that would later serve as the background context for learning a third category. The categories were designed so that the two groups would learn different concepts using the same learned features. Both groups learned that the feature *x* characterized the first category *X*. The two groups differed on the nature of the second category. The first group was exposed to an *XY* category defined by the *x* feature adjacent to the *y* feature. The category of the second group was defined only by the *y* feature. Subjects then learned a third *XYZ* category defined by adjacent *x, y,* and *z* components. Subjects' encoding of the new category was tested with a sorting task and a same–different speeded judgment task. It was found that the second group of subjects, but not the first group, distinguished *XY* stimuli from *XYZ* stimuli. These results confirmed the hypothesis that different histories of categorization generate different feature spaces to encode similarities and contrasts between objects.

2.3. Two types of concept learning

The experiment described above indicates that a history of categorization can trigger different concept learning mechanisms. By the time the third concept is to be acquired, subjects of the second group have the necessary features *x* and *y* to identify the third category; subjects of the first group must create a third, novel feature, *z,* in order to identify the third category.

In the concept learning space fixed by *x* and *y,* Group 2 subjects represent *XYZ* as a combination of the two previously acquired features. This particular encoding illustrates what we call "fixed space learning," the familiar diagnosticity-driven learning that Gibson (1991), Lawrence (1949), and concept learning researchers have discussed. However, the combination of *x* and *y* already represents the second category of Group 1, so subjects must develop a new feature, *z,* to distinguish the third category (see Rodet & Schyns 1994). We call this encoding "flexible space learning," to emphasize the expansion of the categorization space to include a new feature or dimension.

2.4. Functional features and primitives

The premise that features are created to subserve categorizations applies to the creation of functional features but is neutral regarding their perceptual realization. For example, the object property "square" could be featurally represented as a concatenation of image pixels, as four line segments, as four corners, as four smaller squares, as two smaller rectangles, as a linear combination of sinusoids, and so forth. In short, there are many possible realizations of a functional feature. We have proposed that properties of an object that become diagnostic for important categorizations can become functional features of a system's repertoire. However, one potential problem that must be addressed is the degree to which these functional features are themselves based upon a (more) primitive set of features. If a primitive set of features can capture all the regularities and categorizations accommodated by the functional features, then the new functional features do not increase the representational capacity of the system. If this is the case, then the hypothesis that feature creation is necessary to allow a system to represent object properties that it was previously incapable of representing cannot be maintained. We will argue accordingly that functional features are not always constructed from a fixed catalog of primitive features.

A set of shape primitives that could ground categorization must satisfy at least three conditions: the primitives must exist prior to experience with the objects they represent, they must be sufficient to represent the entire set of representable objects, and they must be able to bootstrap complex recognition systems. Ultimately, there are two ways of conceptualizing these primitive features, each with its own problem. Either primitives are fine-grained and relatively unstructured, or they already represent complex structures of the environment.

2.4.1. Unstructured primitives. According to the *unstructured* approach, if one takes sufficiently fine-grained primitives (e.g., very small line segments, or even pixels) together with powerful combination rules, diagnostic compositions of the primitives could represent complex properties of objects. However, functionally important object regularities (symmetry, serif, beauty, etc.) are often not captured by simple pixel-based features. It is unlikely that systems that hypothesize object properties such as symmetry as a primitive of object recognition (Gibson 1969) can explain them by commonalties at the pixel-level (but see Barlow 1980; Barlow & Reeves 1979). Moreover, as is discussed below in the section on formal models of feature extraction, it is not practically feasible (although it is logically possible) to extract relevant categorization features from pixel-based (or similarly unstructured) representations of the input.

2.4.2. Structured primitives. According to another approach to primitives, the catalog includes more complex primitives, such as larger curves, corners, squares, circles, triangles, or even three-dimensional shapes such as cones and cylinders (see, e.g., Biederman 1987; Garner 1974; Treisman & Gelade 1980). Complex (rather than simple) primitives would already mirror important structures of the visual environment and could therefore account for complex recognition by initially segmenting the visual environment into useful primitives for recognition. However, such preformed recognition systems are blind to structures that are not represented as primitives and that are not compositions of simpler primitives.

To illustrate, in Fisher's (1986) influential model of letter recognition (cited in Czerwinski et al. 1992), a capital "A" is identified by composing three primitives (two diagonal bars and a horizontal one). Clearly, diagonal and horizontal bars were selected as primitives with the task of letter categorization in mind; the same primitives would be particularly clumsy in categorizing varieties of ellipses. One might imagine adding a second subset of primitives for distinguishing ellipses. However, any large-scale, highly structured set of primitives is bound to be too coarse to detect (and internally represent) all of the distinctions that might be required by different categories of objects.

2.4.3. Interactions between choice of primitives and task constraints. Task constraints almost always influence the primitives that investigators import into their componential theories of recognition. In our view, the task of the subject creating new functional features for categorization is not substantially different from the task of the scientist creating a componential theory of recognition: both must produce a catalogue of features that are defined by their role in recognition. If the investigators want to posit a complete fixed set of primitives, they must envision all possible recognition problems before conceiving of the features that would solve them. So, the envisioned set of tasks influences the primitives of recognition that will be selected by a theory of object recognition. Similarly, the particular categorization tasks confronting individuals influence the units of representation that they will adopt. Thus, rather than drawing a correspondence between a particular theory of object recognition (with its static primitives) and an individual's object recognition capabilities, the proper correspondence may be between the individual and the metatheoretic search for a proper object recognition theory.

2.5. Functions, perceptions, and their interactions

One constraint on the creation of features is their usefulness for categorization. Our hypothesis that there are functionally determined features does not imply that physiological or sensory facts are unimportant for defining the feature repertoires. Features are also based on general perceptual constraints, such as contiguity, topological cohesion, changes of curvature, and perceptual salience. In many cases, these constraints are not a catalogue of shape primitives, but the constraints nonetheless exert strong pressures to create certain features. To illustrate, Hoffman and Richards (1984) have proposed that objects are segmented by creating parts with endpoints that are local minima of principal curvature. Instead of assuming that objects are segmented into primitive shapes, the authors suggest that a particular patch of shape will be identified as a part because it lies between two points of negative minima of curvature, not because it matches a primitive element. This approach does not limit possible shape features to the compositions of a catalog of primitives. Instead, as a sheet adjusts to the surface on which it is thrown, new features can be acquired to mirror the shapes lying between the segmentations suggested by the minima rule. Hoffman and Richards's constraint on object segmentation illustrates that the structures required for organizing complex representations are not necessarily structured primitives. Instead, general shape-processing constraints can produce segmentations that interact with structuring principles. As Hoffman and Richards (1984, p. 77) state it, "a boundary-based scheme, then, is to be preferred over a primitive-based scheme because of its greater versatility."

A very interesting aspect of Hoffman and Richards's proposal as it applies to the creation of new shape features is that it allows the feature repertoire to mirror partially the shapes that categorizers experience in their environment. This presents new challenges for effective procedures of feature creation. It is conceivable, even desirable, that several distinctive methods are used for developing features, depending on the idiosyncrasies of different object classes. For example, smooth objects such as faces could be parsed into their relevant component features using elastic 3D templates (see, e.g., Hinton et al. 1992). These templates would behave as elastic masks whose parameters would adjust to shape variations within the class. At the time of this writing, there is no agreement on the features, or feature configurations, these masks would have. Class-specific variations (e.g., learning to categorize Caucasian faces) would result in class-specific features that would not be directly applicable to the shape variations of other classes (e.g., Asian faces). Mismatches between expected shapes, and expected shape variations, could give rise to the "other race effect" in which people perceive faces of their own race with greater facility than those of another race (Brigham 1986).

Whereas face stimuli are mostly smooth, many man-made objects are discontinuous. This imposes different biases on the eventual elastic templates and also segmentation constraints other than Hoffman and Richards's minima rule, which operates on continuous surfaces. The templates could be biased to "break" at sharp discontinuities of the surfaces, if a categorization required such a break. Such templates could progressively evolve into a repertoire whose asymptotic state could resemble Biederman's (1987) geons, if they were exposed to many man-made object categories. The extraction of 2D shape features could also require distinct mechanisms and representations. For example, 2D patterns (letters, numbers, textures, etc.) could use feature creation mechanisms based on "growth" (see, e.g., Marr 1982; Ullman 1984). Small 2D patches could locally grow from the interior of a 2D pattern until boundary edges stop the growth. New shapes could then be learned from correlations across category exemplars. To illustrate, consider a simple example of this process (adapted from Schyns & Murphy 1994). Object 1 is a 2D pattern in which arrows show the negative minima that are perceptual indicators of its parts (see Fig. 2). Consider that target and Object 1 (or Object 2) form a category. If a 2D contiguous patch is grown in target, its intersection with the patch grown in Object 1 will identify a part feature (indicated by dashed lines in Fig. 2). A different feature would result from the intersection of target with Object 2.

In short, we are arguing that different object categories are likely to prompt the acquisition of different types of features. These different categories are likely to necessitate differently biased mechanisms. Perceptual biases should facilitate the extraction of features in the objects considered (e.g., smooth vs. discontinuous; 2D vs. 3D). Categorical biases should tune the features for the categorizations performed. The examples discussed suggest the possibility

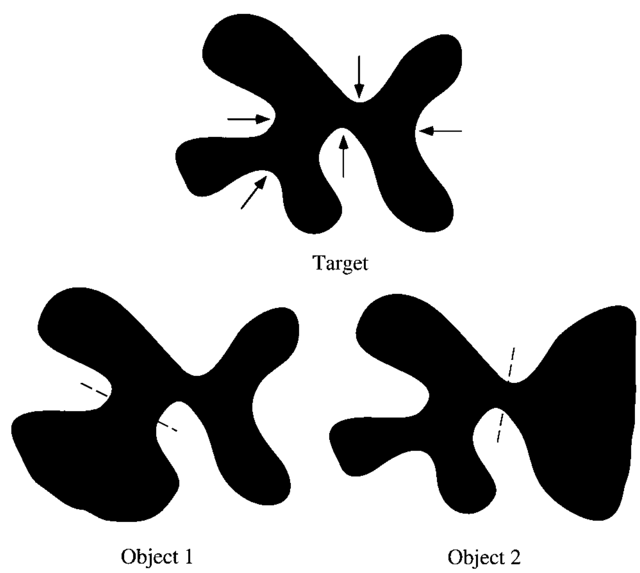

Figure 2. Possible interaction between perceptual and functional constraints in learning new features of object representation. The arrows in the target object indicate perceptual constraints on its segmentation. The Target and Object 1 (or Object 2) constitute a category. The dashed lines on the bottom objects illustrate that the shape features extracted on the target also depend on its category membership. (Adapted from Schyns & Murphy 1994.)

of creating such features, but they do not provide detailed realizations. It will be a difficult (but necessary) task to extract class-specific perceptual biases to build task-specific feature extraction mechanisms. We will come back to this point when we discuss formal mechanisms of feature extraction.

Both functional (categorical) and perceptual constraints determine what features will be created. We see these constraints as mutually interactive rather than strictly sequential (see also Wisniewski & Medin 1994). We might envision a system that first created a set of candidate features by applying perceptual constraints and then selected the new feature from this set of candidates by applying functional constraints. Such a system, however, would suffer from several problems. First, in many cases, an implausibly large number of candidates would have to be considered because objects are underdetermined by perceptual constraints (e.g., a 2D object silhouette with 20 bumps on it would have 380 possible parsings even if only contiguous segments were considered). If functional constraints are considered only secondarily, then processing will be inefficient in that too many candidate features that are not potentially useful will be considered; the constraining role of functionality would not be fully exploited. Second, if strong perceptual constraints are applied (e.g., shape primitives), then the relevant feature will often fail to be in the set of candidates. Third, there is substantial evidence that the functionality of a feature influences relatively low-level perceptual processing (Algom 1992; Goldstone 1994a; 1995; Oliva & Schyns 1995; Rodet & Schyns 1994; Schyns & Murphy 1991; 1994; Schyns & Rodet 1997). The cumulative effect of this evidence makes it unlikely that functionality is considered only after perceptual processing has been completed.

Whereas we admit the intrinsic futility of searching for the boundary between perception and conception, we believe it is useful to describe a *continuum* from the perceptual to the conceptual. What varies along this continuum is how much and what sort of processing has been done to the input. Specifying exactly where experiential and categorical pressures influence processing along the perception–conception continuum is a real, although highly empirical, problem. One apparently fruitful approach to specifying how early an influence conceptual factors have is to identify influences on other processes. Thus, there is evidence that conceptual factors (knowledge of categories and attitudes) not only influence physical and immediate color judgments (Delk & Fillenbaum 1965; Goldstone 1995) but also exert an influence *before* the perceptual stage that creates color afterimages has completed its processing (Moscovici & Personnaz 1991). Similarly, there is evidence that conceptual factors related to one's knowledge of object categories exert an influence before the processing stage that produces figure–ground segregation (Peterson & Gibson 1994).

Another approach to specifying locations of influence on a perceptual–conceptual continuum is to observe the time course for the use of particular types of information. For example, on the basis of priming evidence, Sekuler et al. (1992) argue that knowledge about what an occluded object would look like were it completed influences perception after as little as 150 milliseconds. In general, there are experimental tools available that can identify when – absolutely and relative to other processes – conceptual factors modify information processing. Although the bulk of the work needed to specify the precise locus of influences has yet to be done, current evidence suggests a surprisingly early contribution of conceptual factors such as background knowledge and learned categories.

2.6. Feature extraction and experimental materials

For reasons of control, many experiments in concept learning have used very simple stimuli varying on clearly demarcated dimensions. Real-world objects often vary along many dimensions, however, and in most cases it is difficult to know what the relevant dimensions are. Although there are excellent reasons for using simple, easily described experimental materials, one major disadvantage with this approach is that it may systematically underestimate the importance of finding an appropriate encoding for the stimuli. It may even be that the traditional use of simple materials produces a bias against finding evidence for feature creation.

Table 1 illustrates some properties of different types of materials used in experiments. The properties listed in the left column characterize many typical stimuli used in concept learning experiments. The properties listed in the right column, "alternative materials," characterize materials that are likely to promote the creation of new features during concept learning. Conceptually, all the properties listed in the left column serve to make task-relevant features easy to isolate and identify. Conversely, the properties listed in the right column make it likely that the relevant features are not originally encoded, but they allow for their derivation.

Alternative materials are typically dense (Goodman 1965) in that there is no limit on the amount of information that can be obtained from the input or the number of interpretations that can be made. Therefore, alternative materials may contain many different levels of intrinsic structure, allowing for the potential relevance of highly diverse feature sets. Many blobby structures can be extracted from, for example, X-ray pictures that are not combina-

Table 1. *Stimuli typically used in concept learning and stimuli likely to give rise to encoding of new features*

Traditional materials	Alternative materials
Properties of dimensions in isolation	
Discrete	Analog/continuous
Symbolic	Subsymbolic
Parts easy to delineate	Parts difficult to delineate
Few features	Large number of potential features
Relevant features are salient	Relevant features are not salient
No emergent properties	Emergent properties
Single level of analysis	Multiple levels of analysis
Large dimension value differences	Small dimension value differences
Properties of dimensions in context	
A priori diagnostic features	A priori nondiagnostic features
Features have constant instantiations	Features are variably instantiated

tions of a priori diagnostic features (except to radiologists). Conversely, traditional materials embody a single level of analysis into features known a priori. The primary level of analysis for alternative materials is subsymbolic, because they are designed to ensure that symbols ("square," "circle," "has legs," etc.) are not easy to assign a priori to the important structures of the stimuli. Stimuli that are likely to be represented in an analog fashion may preserve topological relations that leave open the possibility of a stimulus reinterpretation if new categorizations require such a reinterpretation. Discrete stimuli do not allow this possibility because their interpretation is often unequivocal and automatic. Figure 3 presents several examples of alternative materials that are used in our experiments on feature creation. Figure 3a shows a Martian rock (Schyns & Murphy 1991; 1994), Figure 3b some doodles (Goldstone, in preparation), Figure 3c some Japanese hiragana characters (Ryner & Goldstone, in preparation), Figure 3d shows from left to right an *XY* and an *X* Martian cell (Rodet & Schyns 1994), Figure 3e, a Martian lobster (Thibaut 1995), and Figure 3f, a Martian landscape (Schyns & Thibaut, in preparation).

The task confronting subjects who are given what we are calling alternative materials is similar to the task confronting the child who must learn such concepts as *dog*, *table,* and *father.* The child must learn the features that make up

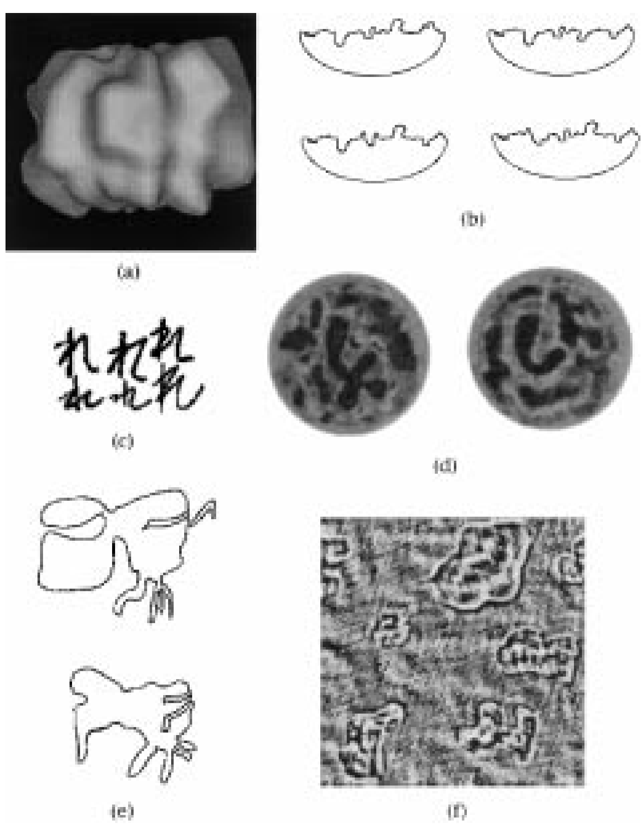

Figure 3. Examples of the alternative materials used in our experiments. Picture *a* shows a Martian rock (Schyns & Murphy 1991; 1994), *b* shows some doodles (Goldstone, in preparation), *c* shows some Japanese hiragana characters (Ryner & Goldstone, in preparation), *d* shows two Martian cells (Rodet & Schyns 1994; Schyns & Rodet, in press), *e* shows a Martian lobster (Thibaut, in preparation), and *f* shows a Martian landscape (Schyns & Thibaut, in preparation).

these objects in addition to learning the proper characterization of the concept. Many formal approaches to categorization explicitly avoid issues of feature representation. Researchers often adopt a stance of "You tell me what the features are, and I will tell you how they are integrated to perform the categorization." Such formal approaches often place no constraints on what may count as a feature. In fact, the lion's share of the work in concept learning seems to be in finding the "right" description space for concept learning.

2.7. Evidence for novel functional features

Novel features are sometimes created and may not be reducible to previously existing features of the system. One version of this claim is certainly false. Novel visual features are certainly reducible to their retinal encodings, and possibly to existing structures at early, lower level representations. Thus, it is a conceptual challenge to characterize a "novel feature." Part of the difficulty is that novelty implies a reference point. At the level of the retina, different encodings of the same object are always novel owing to differences in the retinal projections of the input. However, conceptual encodings of this object are much more stable. Functional, high-level features presumably supply the basis for this stability in the cognitive architecture. The question thus becomes, When is a functional feature novel?

Functionally, a feature may be novel simply because it encodes a categorization that was not performed previously. Our conception of functionality is more constrained than this, however, referring to the synthesis of new elements from raw data. There are two difficulties with the latter variety of novelty. The first difficulty is preexistence: How do we show empirically that a "created" functional feature did not exist prior to the categorization problem? The second difficulty is reduction. How can we ensure that a "created" functional feature does not result from the combination of preexisting functional features?

An ideal empirical test of preexistence would demonstrate that a functional feature, f_x, not initially present in the feature repertoire becomes a member of the set as a result of learning a new categorization. The absence of f_x from the initial repertoire, together with successful categorizations of the new objects, would suggest that f_x was created (instead of merely weighted for its diagnosticity), assuming that f_x is required to perform the categorization. However, empirical evidence for the absence versus presence of f_x is limited to a behavioral manifestation of the new feature (e.g., in a transfer or priming task). Unfortunately, a nonexistent feature is behaviorally equivalent to an existing feature with an "attentional weight" of 0. This makes it difficult to tease apart feature weighting from feature creation based on simple, direct tests of the existence of a feature in memory. Evidence of feature creation is, hence, necessarily indirect, testing the implications of foundational assumptions of fixed feature theories. Two of these assumptions are (1) that objects are characterized by a prespecified, fixed, unambiguous, and nondecomposable set of features and (2) that learning always selects, combines, and weighs the fixed features that tend to characterize categories. An important implication of these two assumptions is that category learning is only strategic. That is, learning weighs features of the fixed set, but it does not change the perceptual analysis and the perceptual appearance of the input.

One way to provide evidence for feature creation would be to show that category learning changes features that participate in the perceptual analysis of identical stimuli. This was the goal of the experiment of Schyns and Rodet (1997) described above. This experiment was controlled so that the features x and y were each diagnostic of one category in the two categorization conditions ($X \rightarrow Y \rightarrow XY$ and $XY \rightarrow X \rightarrow Y$). Hence, they should in principle elicit identical featural analysis and identical perceptions of the same category exemplars; that is, subjects in the two categorization conditions should equally see XY exemplars as feature conjunctions. However, the outcome was mutually exclusive perceptions of XY stimuli (a conjunctive and a configural perception), making a feature weighting interpretation of these data difficult to justify. Feature creation is preferable to feature weighting if category learning induces a mutually exclusive perceptual analysis of an objectively identical object property when the experimental design would predict identical perceptual analysis if the subjects used fixed features.

The other problem of feature reduction is comparatively simpler to address empirically. In principle, if a functional feature is the combination of two or more other features, these other features would become active each time the new feature was presented. However, priming tests on these subfeatures would indicate whether or not they participated in the perceptual encoding of the new feature.

It is always difficult to refute a feature weighting interpretation of categorization results. Part of the reason is that feature weighting is difficult to refute when it is used *a posteriori* to interpret patterns of data. Feature weighting is a form of curve fitting with free parameters (the weights assigned to features). Feature weighting therefore covers not one model but a potential infinity of models of categorization and can potentially accommodate any pattern of experimental data if its features are not prespecified. Attempting to explain features through the history of categorization allows the theorist to ask an important question: What counts as a feature? Most concept learning programs do not address this, but they nonetheless call for new features in different situations. We accept the need to generate different feature sets for different tasks, but we would like the theorist to explain how the features come to be generated, instead of simply positing their existence.

2.8. Advantages of new feature learning

A system that allows for the creation of new features during concept learning offers several advantages over fixed feature set approaches. First, the most basic advantage, alluded to earlier, is that an ability to acquire new features allows flexible but constrained features. Unlike purely formal models of similarity and categorization, our approach places constraints on what can count as features: features will be incorporated into a system to the extent that they distinguish between object categories; features should not be limited to the finite set of a priori features designed by a particular researcher for a particular domain.

Second, a learned set may be equivalent to, but not limited to, other proposed fixed feature sets. Fixed feature sets are motivated by design considerations and psychological evidence. For example, Biederman (1987) suggests that evidence in favor of geons as primitive features comes from studies that delete line segments from objects. When line segments are deleted in a way that does not allow geons to be recovered, object recognition is particularly impaired. However, to the extent that geons are useful features for object categorization, it is reasonable to suppose that they might be generated from functional constraints applied to simpler building blocks, such as line segments, or corners, or surfaces. Consequently, evidence in favor of a particular set of features does not entail that the set of features is hard wired.

Third, a learned set permits a near-optimal fit between categorization demands and the expressive repertoire. New features are created to represent new categorical commonalties or contrasts and can be optimally adjusted in number to a wide variety of task demands (e.g., expert categorizations and subcategorizations). To the extent that each new feature accommodates at least the categorization for which the feature was created, the repertoire should be free of useless features. A fixed feature approach is necessarily much less parsimonious: many spurious features must exist in the feature repertoire for the individual to foresee new categorizations. Moreover, most features of the fixed set would never be used; they would keep waiting for their "Godot category." Fixed features necessarily have suboptimal fit outside the scope of the stimuli they were designed to represent.

Fourth, a flexible set of features tuned to specific categorizations reduces the necessity of complex categorization rules. To illustrate that good representations often carry most of the burden of categorization, consider the XOR problem in learning theory. XOR is a binary function categorizing the pairs (0, 0) and (1, 1) as members of the "0" category and the pairs (1, 0) and (0, 1) as members of the "1" category. Categorization rules that separate the "0" from the "1" category are complex and nonlinear, because no linear solution (a straight line) achieves the separation. Complex learning problems often become simpler with better representations. Add another number as a third input to XOR, which is 1 whenever the two input numbers are (1, 1) or (0, 0), and 0 otherwise. A simple recoding simplifies the problem: there is now a linear solution. Although XOR is only a simple formal problem, it nonetheless illustrates the general point that carefully crafted representations often reduce the complexity of categorization processes.

Concept learning theories have frequently stressed the importance of learning categories by discovering complex rules that integrate several distinct stimulus features (Bruner et al. 1956; Nosofsky et al. 1994). Concept learning certainly does sometimes require such integration. However, these problems have effortful, strategic solutions. They are rather unnatural; people are not particularly adept at explicitly combining psychologically separated sources of information. Our alternative is that new categorizations can be based on relatively few, specially tailored features.

Fifth, in the flexible feature approach, categorizations can induce a *decomposition* of features into subfeatures. Consider the contrast between glasses and cans. Early in conceptual development, these objects may be indistinguishable because their memory representation corresponds to a single, undifferentiated feature. Now, assume that the organism must distinguish between these objects. This can be achieved by decomposing the undifferentiated feature into two specific features tailored to glasses and cans.

The acquisition of a new feature that segments an initially undifferentiated, unitary feature could account for conceptual differentiation, for example, the basic to subordinate shift (Tanaka & Taylor 1991), the narrowing of children's lexical categories (Chapman et al. 1986), and the construction of conceptual hierarchies (Schyns & Murphy 1994). Classical accounts of concept learning distinguish between features and concepts (which are combinations of features). However, there is little principled distinction between these constructs. Cars may be usefully represented with features such as wheels, but wheels are themselves concepts, which may be decomposable into features (Schyns & Murphy 1994). Even features such as color, which may appear unitary and unstructured, can be decomposed into subunits (hue, saturation, and brightness) under certain conditions (Foard & Kemler Nelson 1984; Goldstone 1994a).

3. Comparison to other approaches

We have argued that the feature space of object representations is often created to reflect the specific categorization requirements of an organism. We described some of the advantages of a feature set grounded in the organism's history of categorization (i.e., the categorizations the organism had to master plus the corrective feedback it received) over the fixed feature sets proposed by many theories of object categorization and recognition. Our proposal for creating new features touches on several issues related to perceptual and conceptual change. The following sections discuss the similarities and contrasts between our proposal for feature creation and feature chunking, new features in constructive induction, developmental constraints on feature extraction, and formal models of feature extraction.

3.1. Chunking and perceptual unitization

Research in the visual search literature has supported perceptual changes similar to the types of changes that we have discussed. Training or automatization effects occur when people actively search for a particular target shape (for example, the letter "A") in a visual array of distracter letters (for example, "M" and "W"). In Fisher's (1986) model of visual search, letters are represented by simple features such as horizontal, vertical, and diagonal line segments. Similarity between features generally makes it more difficult to find, for example, an "A" among "W"s than an "A" among "M"s; "A" and "W" share two diagonal bars but "A" and "M" have no common feature. However, even when featural descriptions are quite similar, extensive training significantly speeds up search times (see, e.g., Fisher 1986).

Czerwinski et al. (1992) suggested that a perceptual change called perceptual unitization could explain training effects in visual search. Perceptual unitization produces perceptual features from a set of more elementary components. These new features speed up visual search because they recode input objects in a more efficient feature repertoire, a repertoire tailored to the specifics of the search task.

Our feature creation theory has both similarities to and differences from unitization and chunking theory. It is similar in that visual search may be framed as a categorization task of distracters and targets. Chunking can then be viewed as a context-dependent process influenced by the contrasts and similarities between targets and distracters. This reformulation of visual search emphasizes functional constraints that the chunking process must satisfy; units will be formed that allow members of the target category to be distinguished from distracters. It also allows specific predictions to be derived. For example, in Fisher's (1986) and Czerwinski et al.'s (1992) models, chunked features could represent any subpart of the capital letters, the subpart that reliably unifies and distinguishes the categories. Perceptual chunking is probably an important mechanism of feature creation. However, we believe that the principles governing chunking cannot be fully understood without the notion of category contrasts and similarities.

The influences of category contrasts and similarities on the segmentation of objects were specifically studied by Pevtzow and Goldstone (1994). Stick figures composed of six lines were categorized in one of two ways. Different arbitrary combinations of three contiguous lines were diagnostic for the different categorizations. After categorization training, subjects participated in part–whole judgments, indicating whether a particular set of three lines (a part) was present in a whole stick figure. Subjects were significantly faster in determining that a part was present in a whole when the part was previously diagnostic during categorization. The part–whole judgment task is arguably the most perceptually based task used by Palmer (1977) to explore the "naturalness" of a way of segmenting an object into parts. Although Palmer's model bases the naturalness of a particular segmentation on properties of the object (e.g., the proximities, similarities, and shapes of the line segments), the results indicate that the subjects' experience also influences how they will segment an object into parts.

The differences between unitization and functional feature creation are mostly consequences of using discrete vs. continuous stimuli. As its name indicates, unitization requires that the stimuli be discretized before being unitized. However, it is frequently difficult to assess exactly what discretization the visual system initially applies to a stimulus before unitization occurs. Czerwinski et al.'s stimuli and Palmer's stick figures are designed to bias processing according to a particular discretization, line segments (however, the authors acknowledge that they can only hope for this segmentation). These stimuli could give the impression that our perceptual systems initially segment the environment into little line segments and then construct complex task-dependent representations by unitization. However, the varieties of recognition tasks we face make it very likely that there is no single scale of representation.

Many psychophysical and computational models are converging on the observation that perception operates simultaneously at multiple spatial scales and that the coarser scales are often sufficient for effective processing of complex pictures (see, e.g., Burt & Adelson 1983; De Valois & De Valois 1990; Marr 1982; Schyns & Oliva 1994; Watt 1987; Witkin 1986). Multiscale representations suggest that the input stimuli are discretized at different scales, possibly using scale-specific feature repertoires. If line segments may serve as the discrete elements at the finer spatial scales (though even here there are serious difficulties), "blobs" or other image measurements are more appropriate for discretizing the coarser scales. A conjunction of high-resolution edges often maps onto a single coarse-scale blob, suggesting that the input signal could initially be parsed into large components that do not result from fine-scale unitizations. Hence, efficient parsings of real-world stimuli could initially operate with the scale-specific primitives closely corresponding to the relevant events of the input signal (see, e.g., Oliva & Schyns 1995). These scale-specific primitives should be adjustable to scale-specific shapes and should therefore be sensitive to task contingencies (Oliva & Schyns 1997). Scale-specific vocabularies could arise by applying our proposal for learning new features to the spatial scales made available by perception.

In summary, although chunking is probably an important mechanism for creating new perceptual features, we think that there are alternatives. Chunking applies only to a priori discretized stimuli, but evidence suggests that stimuli are not unequivocally discretized into their smallest structures (or for that matter into a single, preferred scale). Large features may be registered without being composed out of smaller features, and small features may sometimes be created by decomposing larger features.

3.2. Constructive induction

The idea of creating new featural descriptions has been a direct concern of a branch of machine learning called *constructive induction* (Matheus 1991; Michalski 1983). In constructive induction, new features are created by applying inductive operators to the existing set of features. For example, objects that belong to a category may originally be described as 74, 78, or 71 cm tall. With the "close interval" operator, a single new feature "any height between 70 and 80 cm" may be created. Generally, the operators that have been considered have been highly symbolic, including logical operators such as "and" and "or," and hierarchical relations between category classes. To give an example of the latter type of operator, a playing card that was originally represented as "diamond" may be recoded as "red" if the system knows that diamonds are red.

Hofstader and his colleagues (French & Hofstadter 1991; Mitchell 1993) have also been concerned with computational systems that create new descriptions for input patterns. For example, Mitchell and Hofstadter's Copycat system, when processing the letter sequence "PPQRR," may develop either the description "P, followed by the series PQR, followed by R" or the description "a Q in the middle, flanked by a pair of Ps on the left and a pair of Rs on the right." The description that emerges will depend on the other developing structures. Copycat creates new descriptions by establishing groups of related letters and by relating these groups.

Wisniewski and Medin (1994) have recently provided evidence that people alter their verbal descriptions of objects to fit the category labels provided (see also Medin et al. 1993). The same figure in a child's drawing may be interpreted as a tie or buttons, depending on how the drawing is labeled. The authors argue that new descriptions are created when links are established between abstract background knowledge (e.g., "creative children should show more detail in their drawings") and concrete object information.

Our proposal for learning new features is consistent with these proposals. Although many of the ideas are similar, our emphasis is different in several respects. We have stressed that relatively raw stimulus properties must be preserved for new features to be created. As was argued above, if distilled, symbolic representations are used to create new features, then there will be severe limitations on what new object features are possible. This is the case with typical constructive induction systems. Although they can produce an infinite number of new features by successive application of inductive operators, the new features are highly constrained by the object interpretation made from the primitive symbolic features. Both the original features and the new features in constructive induction algorithms are discrete symbols that are the product of an object interpretation process. Far greater flexibility in feature creation can be achieved by beginning with object representations in terms of raw features that have not undergone interpretation. The representation should be raw enough so that both symbolic interpretations of an "X" ("two crossing diagonal lines" and "a 'v' and an upside-down 'v' just touching") can be generated (McGraw et al. 1994). Harnad (1990) has made a similar point with respect to the need for grounding symbols by means of representations that are nonsymbolic.

By stressing the importance of early perceptual representations that implicitly preserve distal object properties, our approach to feature creation also stresses perceptual constraints on feature extraction. Whereas constructive induction techniques can create arbitrarily complex features, features that are generated by humans are constrained by perceptual factors such as topology, spatial proximity, and global coherence. Thus, features that are generated by standard constructive induction techniques may be improperly constrained in opposing ways. They can be too constrained by the initial symbolic representations, and they cannot be sufficiently constrained by properties of our actual perceptual systems.

Another difference is that we have stressed the perceptual changes accompanying feature creation. With standard feature creation techniques, new features are added to the system's repertoire, but there is little reason to suggest that the new features alter the appearance of the described objects. Rather, they alter the properties that will be inferred about the objects. There is a difference in immediacy between *seeing* and *inferring* that an object might be expressed in terms of a particular feature. The psychological evidence that we have reviewed suggests that perceptions of the immediate appearance of objects (e.g., their discriminability and apparent organization) are altered by experience. Mitchell and Hofstader's letter series may provide an intermediate case (see also Chalmers et al. 1992). When people interpret "PPQRR" in a particular way, it may be a cognitive inference, an immediate perceptual phenomenon, or something in between. The same ambiguity seems to exist for the high-level features (e.g., forks, traps, and pawn support structures) that are used by chess experts but not by novices (De Groot 1965).

In summary, work in constructive induction is certainly relevant to the present theory of feature creation. Our approach differs from much of this work in focusing on the perceptual constraints and consequences of feature creation, and the importance of beginning with relatively raw object representations for developing novel interpretations of an object.

3.3. Developmental constraints on object feature extraction

There are in principle an infinite number of ways to represent real-world objects with features. This poses a serious problem for developmental psychologists who must explain how children acquire a particular featural object description from a limited data set. Similarly, in acquiring a new word meaning, children are "faced with an infinite set of possibilities about what a novel word might mean" (Markman 1995, p. 199; see also Jones & Smith 1993; Landau 1994; Markman 1989; Quine 1960). To reduce the indeterminacy of featural representations, it has been proposed that young learners come equipped with biases toward particular properties of stimuli that increase the speed and accuracy of learning (Eimas 1994; Landau 1994; Markman 1995). These biases are of two sorts: theories and beliefs about objects in the real world, and perceptual structures and processes. We discuss how these biases constrain the development of functional features, and we argue that they must be supplemented by categorization constraints.

3.3.1. The role of theories in object parsing. According to an influential account of conceptual development, new features and concepts are direct consequences of the development of theories, that is, naive mental explanations of phenomena (Carey 1985; Gelman 1988; Keil 1989; Murphy & Medin 1985). Perceptual features (e.g., *body_shape, length_of_legs, number_of_legs*) lie at the periphery of concepts, whereas our theories place the causes of category membership (e.g., a genetic code) at the core of conceptual organization. It has been suggested that the conceptual core exists prior to experience with the world and that it could bias the features young infants notice in objects (Carey 1985; Spelke 1994).

There are two different views regarding the development of theories, discontinuous and continuous. In the discontinuous view, the conceptual core develops through a differentiation process: new explanatory constructs (concepts and features) result from the differentiation of the existing constructs of an earlier theory (Carey 1991). Children's theories may be incommensurable with corresponding theories in adults (Carey 1985; 1991; Keil 1989; Smith et al. 1985).

Spelke (1994) suggests that, contrary to the discontinuous view of theory development, there is continuity with respect to theories used during conceptual development. For Spelke, there is an innate, constant core at the center of the (intuitive, naive) knowledge later used by older children and adults. Spelke argues that the constant core consists of general constraints that govern the way children perceive and reason about objects in different domains. As Spelke (1994, p. 439) puts it, "learning systems require perceptual systems that parse the world appropriately." Spelke suggests that, among other constraints, an innate cohesion principle biases children to group parts that move together into single objects (see also Eimas, 1994, for a related point of view). This principle could facilitate the parsing of objects from their background and could bootstrap category learning.

In a continuous or discontinuous view of development, innate knowledge is important because it reduces the indeterminacy of featural descriptions to those dictated by preexisting theories. In general, however, there is a conceptual difficulty with the idea that (innate) theoretical knowledge constrains perceptual information: going from theories to predict perceptual data is underconstrained. To illustrate, if a categorizer is instructed that a set of objects with an unknown complex structure is a set of hammers, an existing theory of *hammer* would list the components representing these objects in memory. However, unless the theory also specifies all possible perceptual appearances of these components, a segmentation procedure would still have difficulties locating the actual parts in a new object: the perceptual realization of the parts depends on the new stimulus itself. This is analogous to the symbol grounding problem (Harnad 1990).

Thibaut (1994) has recently investigated mappings of theories on perceptual features. In a feature circling task, subjects were instructed to parse the stimuli of a category of unfamiliar objects that displayed the same overall shape and structure (see Fig. 3e). All subjects were given a category name so that the corresponding general knowledge could assist their segmentations. When asked to name the segmented parts, subjects did not use the same name (e.g., the same part could be called "head," "leg," or "body" by different subjects). Thus, even when a theory provides a listing of the parts to be searched, the assignment of each part to a perceptual structure is not completely constrained by theories (Thibaut & Schyns 1995).

3.3.2. The early role of perception in object parsing. Theories are one source of constraints to reduce the perceptual indeterminacy of stimuli. However, it has recently been suggested that perception also biases children's predispositions toward objects. Experimental evidence has revealed that category inductions are guided by a bias for the shape of objects (see Jones & Smith 1993, and Landau 1994, for reviews of the relevant data). In a typical design, children are presented with a novel three-dimensional object with a novel name (a count noun) and are then asked which objects (of a set of objects that have, or do not have, the same shape, texture, and size) should be called by the same name. Their performance is compared with that of children who are simply asked to select objects that are like the novel object, with no name provided. Converging evidence suggests that children generalize from object names on the basis of shape and neglect large differences in other object properties. This bias appears to develop until the age of 2 years; later, the shape bias predominates only when children are given a count noun (see Jones & Smith 1993).

The shape bias is intended to reduce some of the indeterminacy of category induction. However, because complex shapes are decomposable into many different sets of components, a bias toward shape is only a first necessary step. Other constraints are required to guide the decomposition of a particular shape into its features. In other words, it remains to be explained how children learn to decompose a set of objects into their relevant object features. Such an explanation of parsing could extend the shape bias to specifying precisely which aspects of shape attract attention (and therefore bias segmentation) at different stages of development. It is conceivable that early biases for shape are later superseded by biases resulting from experience with particular object categories. As was argued above, segmentation routines for different categories of geometrical objects (e.g., continuous vs. discontinuous surfaces) could develop and help in making the fine segmentations required by conceptual expertise.

Thibaut (1995) explored the development of segmentation skills in different age groups. Adults and children aged

4 and 6 years were instructed to learn a category of unknown stimuli and were later tested on the parsing of its exemplars. The stimuli shared a global shape and were composed of a common set of shape features that varied slightly across exemplars (see Fig. 3e). Children's parsings were highly inconsistent compared to those of adults. For example, although component parts kept the same relative locations across exemplars, children's parsings often violated topological coherence. They changed the location of the same part across exemplars, and the number of segmentations was not constant across stimuli. Together, these inconsistencies stress that, when children attend to shape, they can be biased toward local similarities between shape aspects at the expense of a consistent integration of shape aspects across instances. Consequently, the new shape features that children isolate could be structurally different from those that adults extract from identical materials.

This has important implications for category learning. Children's biases toward locally salient properties could impede, or even prevent, their learning of new categories, when these are defined by features that are comparatively less salient. Recent evidence from Thibaut (1995) showed that 6-year-old children could not learn a simple categorization (a first category defined by the perceptual cue "a-group-of-three-legs-plus-one" and a second category defined by "two-groups-of-two-legs") when the sizes and orientations of the legs that were irrelevant for categorization varied across exemplars. However, children of the same age experiencing the categories without variations across exemplars had no difficulty learning the categories. These results emphasize the interaction between the development of a feature repertoire and specific perceptual biases. Over the course of conceptual development, children must learn to neglect irrelevant features of the stimuli when they learn new categorizations. The processing differences that could explain the determinants of children's object segmentations should be an important area of future research.

In summary, we have presented theories and perceptual biases as possible predispositions of children toward specific object properties. These biases were not sufficiently specific to predict the actual segmentation of an object belonging to a category. The structure of the categorization problem itself could be an important constraint on the featural descriptions of objects, but it remains to be explained exactly how young children utilize this structure to discover relevant object features.

3.4. Formal models of feature extraction

The understanding of the mechanisms and biases constraining the discovery of relevant structures in data is not the province only of developmental psychologists. For decades, mathematicians and statisticians have been confronted with the issue of structure, as the following quotation from a textbook on morphology illustrates (Serra 1982, pp. 57–58):

> The universe of all possible object shapes is vast, even when it is reduced to equivalence classes. . . . There is therefore a huge offering of potential structuring elements. Thus, analyzing the same object X by two dissimilar structuring elements results in two profoundly different pieces of information on its geometric structure. . . . Only the interaction of X with structuring element B has an objective meaning.

Formal techniques for finding relevant structures in data could provide useful analogies for theories of feature creation.

Mathematically, an object is often expressed as an n-dimensional feature vector. Each component of the vector encodes the presence vs. absence or the values of the n attributes describing the object (e.g., its parts and their shapes, colors, and textures). Geometrically, different points in n-dimensional space encode different objects, and categories of similar objects form clouds of points. There are many ways to encode objects, ranging from the raw pixel intensities of digitized pictures to sophisticated properties that are known to be diagnostic for classification, e.g., *number_of_legs, has_wings, has_fur, has_feathers,* and *hibernates.* Although the latter representation would describe animals in an appropriate feature space, pixel arrays would require extensive processing before diagnostic properties were captured. Our proposal for functional feature creation concerns the extraction of new structures from perceptual data. How could *has_feathers* be discovered from a training set of pixel arrays or similarly unstructured representations?

3.4.1. Properties of high-dimensional spaces and the bias–variance dilemma. Many models of concept learning have successfully shown that category representations can be learned from exemplars when they are composed of a small, prespecified feature set (see, e.g., Gluck & Bower 1988; Krushke 1992; Rumelhart et al. 1986; Widrow & Hoff 1960); the task is not to discover the feature set from high-dimensional raw data. However, it could be argued that the discovery of features from such high-dimensional spaces is not substantially different from standard mechanisms of category learning. Both concern the extraction of task-dependent invariants. Standard concept learning models operating in low-dimensional spaces could simply be scaled up to operate in high-dimensional spaces.

One of the problems with this idea is that high-dimensional spaces are mostly empty. To illustrate, imagine discretizing a line, a squared plane, a cube, and a hypercube with tiles of equal size (e.g., 10 tiles per side). There is a geometrical increase (in this example, 10^1, 10^2, 10^3, 10^4) in the number of tiles that cover the objects. If each tile is represented by an n-dimensional data point, the example shows that one needs approximately 10^n tiles to cover an n-dimensional space. If the input distribution varies along many degrees of freedom, a learning problem in high-dimensional space may require an unrealistically large training set to reveal robust features, even if an asymptotic solution exists in principle.

This *curse of dimensionality* (Bellman 1961) imposes severe limitations on the idea of directly applying simple supervised categorization models to discover perceptual features. Typical concept learning models learn category decision boundaries from a set of pairings of exemplars and their respective category labels. Formally, this consists of finding a function, f, which successfully approximates the desired category name y from an input x. Often, f is chosen to minimize a cost function. Popular concept learning networks minimize the sum of the square of the error between the estimated and the desired category labels (see, e.g., Rumelhart et al. 1986; Widrow & Hoff 1960).

Generally speaking, "error-based" categorization models such as back-propagation are *nonparametric statistical* models (Geman et al. 1992). They are nonparametric because the networks are not biased to particular classes of solutions. Instead, the architectures are unbiased such that they discover structures from data flexibly. Mathematical analysis has shown that the error term (specifically, the expected mean square error) of these networks can be algebraically decomposed into a bias and a variance term (see Geman et al. 1992, pp. 9–10). These two terms summarize the *bias–variance dilemma* (Geman et al. 1992). Networks make a bias error when they are dedicated to a class of solutions that is not appropriate for the categorizations at hand. Such networks may be too rigid, and flexibility (low bias) would be needed to extract task-specific features. However, low bias comes at the cost of high variance, the second component of the error (where *variance* means the discrepancy between the correct categorization and the categorization of the network). There is high variance because a flexible system is too sensitive to the data: it learns many idiosyncrasies of the exemplars (e.g., differences in lighting conditions, rotation in depth, translation in the plane) before learning the invariants of a category. Consequently, experience with many exemplars is necessary for the network to "forget" idiosyncrasies and learn relevant abstractions. Only with great experience is the system able to categorize accurately (keep the variance low). The curse of dimensionality is such that unbiased machines designed to discover many types of new perceptual features flexibly will often require implausibly large training sets to achieve good categorizations. Note that this problem does not greatly affect fixed feature models, which usually operate in smaller spaces for which sufficient exemplars can be generated. The bias–variance dilemma addresses practical computability, not principled limitations.

An ideally flexible system should be constructed so as to keep bias and variance low, using a reasonable training set. The bias–variance dilemma is somewhat analogous to the contrast between structured and unstructured features discussed above. By analogy, fixed sets of structured features make it difficult to learn new categorizations (and therefore raise the bias error). In contrast, unstructured systems will tend to capture irrelevant aspects of the input set that have little relation to the actual basis of categorization (and therefore raise the variance).

3.4.2. Dimensionality reduction. Complex supervised categorization problems in high- dimensional spaces would be simplified were it possible to reduce the dimensionality of the input. Several linear and nonlinear dimensionality-reduction techniques have been designed to achieve this goal. Underlying dimensionality reduction is the idea that information processing is divided into two distinct stages. A first stage constructs a representation of the environment and a second stage uses this representation for higher level cognition, such as categorization and object recognition. It is hoped that the constructed representation in a smaller dimensional space is more useful than the raw input representation.

To illustrate, consider the popular technique called *principal components analysis* (PCA). If redundancies exist in the input data, there should be fewer sources of variation than there are dimensions (i.e., $p << n$). PCA finds the first k orthogonal directions of highest variation in a data set. If each input vector of a high-dimensional space is recoded in terms of a linear combination of the first k sources of variation, the intrinsic structure of the data will be preserved to a first approximation (see Oja 1982; Sanger 1989). In general, however, the featural interpretation of principal components is often difficult because orthogonal directions of highest variance have little connection to the best projections for categorization. That is, there are no *psychological* constraints on the principal components. Principal components need not be spatially or topologically coherent (perceptual constraints) or summarized by a single explanation (conceptual constraints).

Other dimensionality-reduction techniques aim at reproducing the intrinsic structure of the input space. Examples of these range from Shepard's (1957) early multidimensional scaling and Sammon's (1969) nonlinear mapping to more recent Kohonen maps (Kohonen 1984) and a promising extension to Kohonen maps called curvilinear component analysis (Demartines & Herault 1997). These algorithms project an n-dimensional space on a smaller p-dimensional space, while keeping most of the information about the organization of the input space. To illustrate, consider two distinct clouds of points forming two categories in a "high"-dimensional space composed of four dimensions ($n = 4$). Assume further that exemplars of the first category are identical on two dimensions, whereas exemplars of the second category have only one dimension in common. The points of the first cloud lie on a plane ($p = 2$), but the points of the other category have 3 degrees of freedom ($p = 3$) and, therefore, are in three-dimensional space. This simple example illustrates that data sets can have local distributions with different intrinsic dimensions of variation (in the example, 2 and 3). Projections of high-dimensional inputs onto lower-dimensional spaces should account for these intrinsic characteristics if they want to preserve the important degrees of freedom of the distribution. Unfortunately, techniques for discovering the intrinsic dimensionality of a data distribution are also plagued by high dimensionality. The number of data points necessary to estimate reliably the structure of a distribution can be enormous if the intrinsic structure is high. Dimensionality-reduction techniques also have to give up generality for biases, at the expense of possibly missing "important" structures in the data. Nevertheless, the existence of low-dimensional somatosensory maps in cortex clearly demonstrates that brain structures are particularly adept at reducing high-dimensional inputs to lower-dimensional representations (see Kaas, 1995, for a review). Furthermore, there is now growing support for the notion that these natural processes of dimensionality reduction are flexible, allowing different types of reorganizations of cortical maps following different forms of sensory deprivation (Kaas 1995).

By analogy with the functional (re)organization of somatosensory maps, we would like the formal definition of "important lower-dimensional structures" to be closer to the categorization task the system has to solve. Recent approaches to dimensionality reduction have incorporated measures of "feature goodness" in the algorithm for determining good dimensions of recoding. For example, Intrator (1992) and Intrator and Gold (1993) discuss a technique in which input data are projected onto dimensions that have many distinct clusters of data points (multimodal distributions). This unsupervised technique is more likely to discover dimensions useful for distinguishing categories under the assumption that different categories produce clusters within the data. Intrator (1992) reports that his technique worked on stimuli with 3,969 and 5,500 dimensions and that few training data were necessary for extracting robust features. This technique and related techniques based on projection pursuit (Friedman & Tukey 1974) provide methods with interesting biases for exploring high-dimensional data spaces.

In the dimensionality-reduction techniques reviewed, the feature extraction stage operates independently of higher level processes; thus, there is no guarantee that the extracted features will be useful for higher level processes (Mozer 1994). The functionality principle suggests that the categorizations being learned should influence the features that are extracted. In other words, top-down information should constrain the search for relevant dimensions/features of categorization. Therefore, we believe that the serial process of (1) projecting high-dimension space onto a new lower-dimension space and then (2) determining categorization with new dimensions will have to be modified such that the second process informs the first (see also Intrator 1993). However, computational considerations make it likely that different aspects of perceptual feature extraction require strong biases that do not trivialize the categorization problem (as fixed features often do) but that are sufficiently constraining to allow the learning of general features from a reasonable number of data points (a similar opinion is defended by Anderson & Rosenfeld 1988; Geman et al. 1992; and Shepard 1989, among others). It is conceivable, for example, that different constraints will be needed to model the categorization of intrinsically different object classes such as faces, man-made vs. natural objects and textures, natural and artificial scenes, and so forth. The empirical study of these psychological constraints and biases should explicitly account for the reported interactions of categorization and perception, even if they significantly complicate the problem.

4. Conclusions

The function of a feature is to detect and internally represent commonalties between members of the same category as well as differences between categories. Either people come equipped with a complete set of features that account for all present and future categorizations, or, working backwards, people sometimes create new features to represent new categorizations. We argue for an approach in which people create features in order to subserve the categorization and representation of objects. We have presented psychological evidence and theoretical arguments for the necessity of flexible features in object categorization theories. Flexible features allow the learning of new but perceptually constrained features when new categorizations must be represented. Thus, given an appropriate history of categorization, a learned set of features may be equivalent, but not limited, to proposed sets of fixed features. As new features are created to detect and represent new categorical contrasts and similarities, a learned set permits an efficient fit between categorization demands and the feature repertoire, which should then be free of useless features. Flexible features are inherently linked to categorization tasks and therefore reduce the need for complex categorization rules by providing efficient representations. In addition, advantages can be accrued by decomposing features into subfeatures, without representing all possible decompositions of a holistic feature a priori. In our view, there is little difference between concepts and features: someone's unitary concept might be someone else's decomposable structure, depending on the individuals' histories of categorization.

Experimental materials are more likely to promote feature creation when they are not designed with a priori diagnostic features, leading to obvious feature decompositions. These alternative materials (see Figs. 2 and 3) (1) do not limit the information that can be obtained from the input, (2) have many distinct intrinsic structures, and (3) are not exhausted by their symbolic descriptions, such as "has-legs," "square," "circle," and so forth. In short, alternative stimuli evoke a representation of their structure in a raw, analog form, in a form allowing for a stimulus reinterpretation if new categorizations require such a reinterpretation.

In our view, two types of category learning should be distinguished. Fixed space category learning occurs when new categorizations are representable with the available feature set. Flexible space category learning occurs when new categorizations are not representable with the available features. Whether fixed or augmented learning occurs depends on the requirements of a particular categorization task. That is, it depends on the featural contrasts and similarities between the new category to be represented and the individual's concepts in memory. Fixed feature approaches face one of two problems when they are confronted with tasks that require new features. If the fixed features are fairly high level and directly useful for categorizations (such as Biederman's geons), then they will have insufficient flexibility to represent all objects that may be relevant for a new task. If the fixed features are small, subsymbolic fragments (such as pixels), then regularities at the level of functional features, regularities that are required to predict categorizations, will not be captured by these primitives.

Flexible features and the perceptual learning they allow have important similarities to, differences from, and implications for various fields of cognitive science. Perceptual unitization similarly implies that recoding proximal stimuli with new features affects the perceptual appearance of the distal object. However, unitization assumes that stimuli are initially analyzed into components before being unitized, whereas evidence suggests that stimuli are not unequivocally discretized into their smallest structures (or, for that matter, into a single, preferred scale). Functional constraints influence the scale of discretization. The field of constructive induction in artificial intelligence is concerned with creating new object descriptions to assist in categorization. In many cases, the new descriptions are simple symbolic transformations of existing symbolic descriptions. Instead, we have stressed the need to create object features from relatively raw, unprocessed, perceptual representations and to create new features by incorporating perceptual rather than purely formal constraints. Developmental biases (both theory based and perceptual) that could constrain feature extraction were also reviewed. We argue that neither the shape bias nor a priori theories are sufficiently constraining to predict the actual perceptual features that are discovered in objects. These features are also provided by the structuring role of learned categories. Formal analogies with the principles we discuss are found in statistical techniques of dimensionality reduction and their network implementations. These techniques also attempt to reduce what is initially a high-dimensional categorization space to a lower-dimensional space representing important features. Supervised learning is closer to the principles we discuss, insofar as it explicitly provides feedback to constrain the search for categorization features. However, it must be properly constrained to be practically feasible. We believe properly constrained dimensionality-reduction techniques (techniques constrained by perceptual and categorical factors) come closest to the principles we discuss.

ACKNOWLEDGMENTS
This work was funded in part by an NSERC grant awarded to Philippe G. Schyns and by National Science Foundation grant SBR-9409232 awarded to Robert L. Goldstone.

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.

Eigenfeatures as intermediate-level representations: The case for PCA models

Hervé Abdi[a,b], Dominique Valentin[a,b], and Betty G. Edelman[b]

[a]*The University of Texas at Dallas, School of Human Development, Richardson, TX 75083-0688;* [b]*Université de Bourgogne, 21004 Dijon Cedex, France.* **herve@utdallas.edu www.utdallas.edu/˜herve; herve.abdi@u-bourgogne.fr www.u-bourgogne.fr/abdi**

Abstract: Eigenfeatures are created by the principal component approach (PCA) used on objects described by a low-level code (i.e., pixels, Gabor jets). We suggest that eigenfeatures act like the flexible features described by Schyns et al. They are particularly suited for face processing and give rise to class-specific effects such as the other-race effect. The PCA approach can be modified to accommodate top-down constraints.

How can the gender, race, or identity of a face be inferred from a digitized picture? We can imagine that the "flexible features" proposed by Schyns et al. to perform this type of task correspond to some intermediate-level features that are progressively extracted from the exemplars of the relevant categories. The problem, however, is to find mechanisms responsible for extracting such features. We propose that for categorization tasks involving high similarity object classes such as faces, the principal component analysis (PCA) model is a good candidate. The applicability of this model to face processing, first suggested in the late 1980s (Abdi 1988; Sirovich & Kirby 1987; Turk & Pentland 1991), is still current (Hancock et al. 1996; O'Toole et al. 1997).

The PCA approach represents faces by their projections on a set of orthogonal features (principal components, eigenvectors, "eigenfaces") epitomizing the statistical structure of the set of faces from which they are extracted. These orthogonal features are ordered according to the amount of variance (or eigenvalue) they explain, and are often referred to as "macro-features" (Anderson & Mozer 1981) or *eigenfeatures* in contrast to the high-level features traditionally used to describe a face (e.g., nose, eyes, mouth). Eigenfeatures are flexible in that they evolve with the faces encountered (Valentin et al. 1996) and depend on the set of faces from which they are extracted (O'Toole et al. 1991).

Eigenfeatures, because they are optimal for the set of faces from which they are extracted, are less efficient for representing faces from a different population and thus generate class-specific effects such as the other-race effect. As an illustration, Figure 1 displays the projection of 160 White faces and 160 Japanese faces on the first four eigenvectors derived from White faces (none of the faces has been used to compute the eigenvectors). The Japanese faces are more similar to each other than the White faces are. This shows that White eigenfeatures give rise to the other-race effect when used for categorizing Japanese faces.

Eigenfeatures can be used to perform higher-level categorization tasks such as face categorization or identification. For example, Abdi et al. (1995) showed that a neural network, trained to classify faces according to their gender, generalizes its learning to new faces better when the faces are represented by eigenfeatures than by arrays of pixel intensities. This superiority of the eigenfeatures over pixel representations suggests that eigenfeatures are semantically relevant.

Semantic relevance has also been demonstrated by O'Toole et al. (1993) and Valentin and Abdi (1996). They showed that different eigenfeatures capture different kind of information. As illustrated by Figure 2, eigenfeatures with large eigenvalues contain information relative to the orientation (e.g., full-face, profile) in addition to the categorical assignment (e.g., gender, race) of the faces. These eigenfeatures are robust and can be estimated from a small set of faces (Valentin et al., in press). In contrast, eigenfeatures with small eigenvalues contain face identity information.

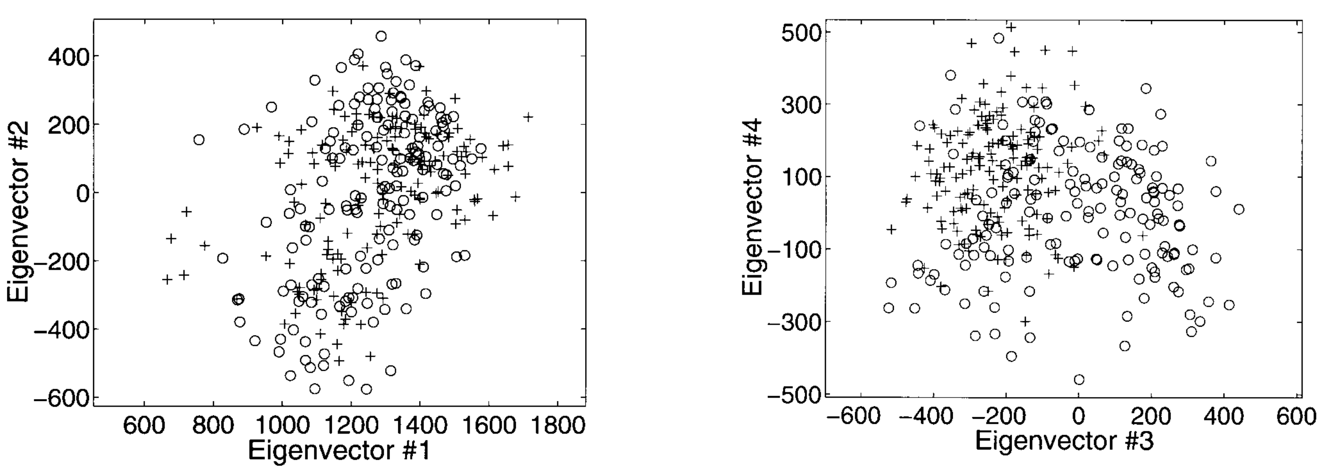

Figure 1 (Abdi et al.). An illustration of the other-race effect. New White faces, (i.e., not learned by the model), denoted by "o," and new Japanese faces, denoted by "+," are projected on eigenfaces obtained with White faces only. The between-faces similarity is larger for other-race faces (i.e., Japanese) than for own-race faces.

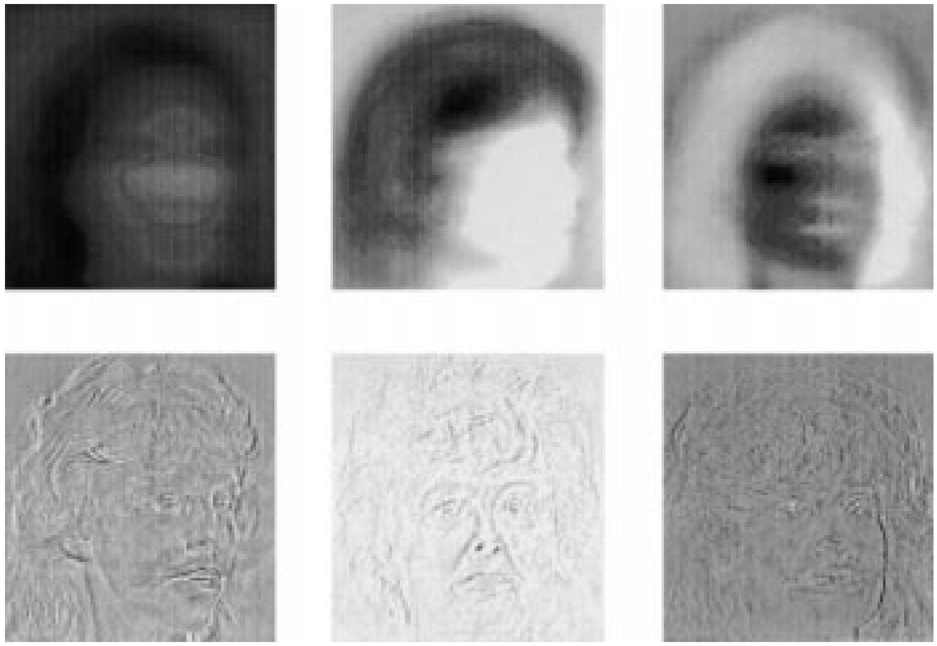

Figure 2 (Abdi et al.). First three (top panel) and last three (bottom panel) eigenfeatures of a set of 40 female faces each represented by 10 views sampling the rotation of the head from full-face to profile with 10-degree steps (from Valentin 1996).

A potential problem for the PCA approach, as noted by Schyns et al., is that feature extraction operates independently of higher level cognitive processes. Using top-down information to constrain the eigenvector representation may offer a solution to this problem. For example, Abdi et al. (1996) have recently derived a generalization of the PCA model that incorporates a priori constraints both at the level of pixel representations and of faces. Using this generalized PCA model, Abdi et al. (1997) showed that constraints on the pixels improve the gender categorization performance of PCA models. The pixels were weighted according to their information content, an idea that is consonant with the proposal by Schyns et al.

In conclusion, eigenfeatures may play the role of the "flexible features" hypothesized by Schyns et al. even though some other mechanisms certainly coexist. A future line of development for PCA models is to incorporate an entry code that is more realistic perceptually or cognitively than pixels such as the output of Gabor filters, or wavelets.

Feature see, feature do

Philip J. Benson

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

Abstract: Physiological evidence predicts a model of concept categorisation that evolves through direct interaction with object feature selection. The requirement stated by Schyns et al. for feature plasticity is supported, but important caveats raise a question about the level at which feature identification can occur. Visual attribute selection for feature creation is likely to be directed by top-down and attentional processes.

Assuming that primary visual cortex (V1) is necessary for object recognition strongly suggests that the geniculostriate pathway is fundamental in bootstrapping the dimensionality reduction process. This is my basis for expanding on the Schyns et al. proposal and their model for representation and categorisation of objects in association cortex.

Ablation of V1 does not disconnect other significant extrastriate areas from all aspects of the visual "stream," areas such as V3a, V5/MT, or regions of the ventral temporal system. These areas continue to communicate "residual" signals via subcortical networks. This explains unconscious visual processing in humans (e.g., "blindsight," reviewed in Stoerig & Cowey 1997). On the other hand, a destriate monkey behaved as if it did not recognise objects in its environment despite good visual acuity (Humphrey & Wieskrantz 1971).

Data from simulations and anatomical projections from the lateral geniculate nuclei to, and laminar architecture within, V1 provide an early feature base using oriented Gabor patches (Olshausen & Field 1996) and, with extension, spatial invariance can be achieved. To date, no other forms of object features or their method of extraction have been described. Bidirectional projections from V1 to the inferior temporal area may pass through surprisingly few stations (V1, V2, V4, posterior and anterior temporal cortex), each incorporating many-to-one feedforward connectivity and successive increases in receptive field size. Each self-organising station consists of overlapping inputs, with the spatial scope of the competing local fields determined by inhibitory interneurons. Each output stage is nonlinear and sigmoidal, leading to inherited binding of the percept's feature repertoire, category similarity, and membership (if known). Visual feature dimension reduction is thus achieved physiologically. One more issue must be considered before the model can be extended.

One must regard the apparent specificity of single cells in association cortex with some caution, but analysis of the information transmitted in firing patterns indicates that competing overlapping ensembles over a distance of up to 2 mm represent a distributed coding of, for example, objects and faces. The removal of ventrolateral frontal cortex impairs discrimination learning; this area is reciprocally connected to inferotemporal association cortex. Nearest-neighbour connectivity with adjacent cortex around the principal sulcus indicates that the ventrolateral frontal area may channel high-order feature and object information relevant to working spatial memory. Such interacting high-level ensembles make for a predictive model of categorisation and interactive feature isolation as follows:

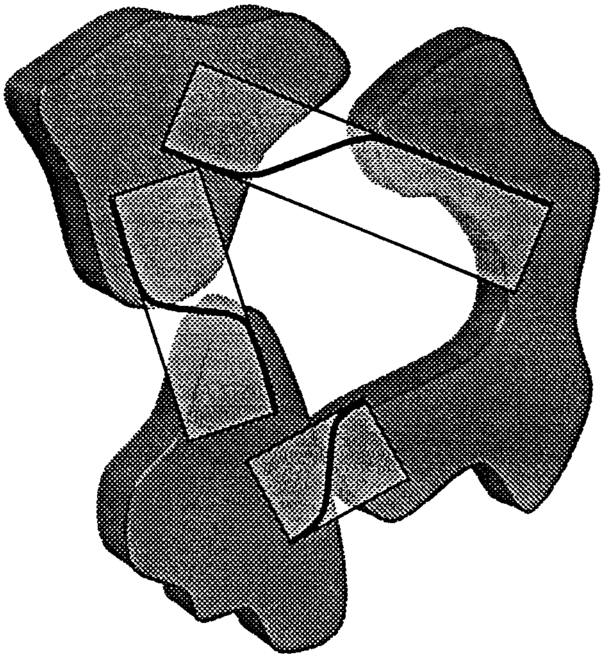

Figure 1 (Benson). The separability of three conceptual object feature spaces can be computed using information theory (cross sections through the convex hyperhulls are illustrated). A high-dimensional trace through a hull is directly comparable to attractor basins or perceptual magnets; attractor minima are one or more integrated features. Categorical perception paradigms (see discussions in Harnad 1987) indicate the concept boundary, and thus the sigmoidal identification functions (shown) and recognition functions are generated and modelled by hull intersection. Successful recruitment of distinctive features modifies the shape of the hulls to subserve encoding. Feature repertoires modelled in this way are predictive of the relative power of component features. They also contain the concept structures necessary to quantify stimulus similarity and energy (how different an exemplar is from its class prototype [Benson 1995a], giving rise to expected discrimination thresholds (signal detection and criterion setting [Triesman et al. 1995]).

For every relevant (detected) feature of a homogeneous class, experience dictates either continuous or discrete measurement. In the former, this leads naturally to a feature vector that includes population sample variance information (variance may be asymmetric about the mean). The identification of a discrete feature immediately enhances categorisability. The union of the sum of all feature vectors represents the range of possible outcomes of the objects' instantiation and thus, mathematically, a convex high-dimensional hull. Informally, a cross section through this hyperhull is a two-dimensional blob from which one may obtain a flavour of multidimensional complexity. The hyperhull model captures object encoding (memory) and decoding (perception) by ensembles.

Categorisation conflicts from overlapping hulls are resolved by reanalysing low-level feature integration and is immediately apparent in the distortion of the hull. Final separation of overlapping hulls signals the utility and validity of the selected feature; repeatedly remembering the hull strengthens the association between the feature repertoire and the category instance and structure. Because the hyperhull embodies object feature variability, "distinctiveness" is inherent to it. Developmental evidence has shown that feature configuration plays an important role in adult life. This is why manipulating veridical and prototypical stimuli is so important in understanding perceptual development (Benson 1995a; 1995b).

Work in progress in our laboratory suggests that the proximal encoding hypothesized to occur in V1 may be complemented by weak projections of subcortical origin from extrastriate to temporal cortex. Thus, although recognition may not be overtly apparent, the performance of a degraded system of feature extraction is desirable and necessary.

Two final comments: (1) How can one model a feature-filled representation of the environment in which one is not aware of higher-order signals, yet one can make accurate responses to them (cf. Crick & Koch 1995)? (2) How could feature plasticity operate in such a system (if at all)? After a critical conditioning period, feature detectors in primary visual cortex are established. Because performance in visual discrimination tasks is driven in a top-down manner by visual experience (Ahissar & Hochstein 1997), this may mean that V4 is a possible candidate for mediating the kinds of featural plasticity described in the target article.

ACKNOWLEDGMENTS

Preparation of this commentary was supported by the U.K. Medical Research Council, the Oxford Medical Research Council and McDonnell-Pew Centres in Cognitive Neuroscience, and the James S. McDonnell Foundation.

A creationist myth: Pragmatic combination not feature creation

Nick Braisby[a] and Bradley Franks[b]

[a]*Department of Psychology, London Guildhall University, London E1 7NT;* [b]*Department of Psychology, London School of Economics, London WC2A 2AE.* **braisby@lgu.ac.uk; b.franks@lse.ac.uk**

Abstract: Schyns et al. argue that flexibility in categorisation implies "feature creation." We argue that this notion is flawed, that flexibility can be explained by combinations over fixed feature sets, and that feature creation would in any case fail to explain categorisation. We suggest that flexibility in categorisation is due to pragmatic factors influencing feature combination, rendering feature creation unnecessary.

If the striking view of features by Schyns et al. is correct, then widely held views of the concepts must be reconsidered. However, the authors have *not* demonstrated that "feature creation" is *necessary* to explain the flexibility of concepts. Moreover, feature creation itself may not be a coherent notion. Instead, we believe that the flexibility of concepts is better explained by pragmatic factors guiding the combinatorial arrangements of a fixed set of semantic attributes.

Schyns et al. argue that fixed feature sets limit the representational (and classificatory) capacity of a conceptual system. They incorrectly claim, however, that "any functionally important difference between objects must be representable as differences in their building blocks" (sect. 1.1, para. 3). This ignores the modes of combination of those building blocks and, were the claim correct, combinations of features would be unable to represent important differences not already encoded in the fixed set. This claim is similar to the (false) assumption that all distinctions expressible in the sentences of a natural language are already lexically encoded. Consequently, feature creation is only *required* in explaining conceptual flexibility if important distinctions are neither expressible by the fixed feature set, nor by *any possible* combination of those features. Schyns et al. fail to address the latter possibility.

Enhancing the expressiveness of a fixed feature set raises an induction problem (cf. Fodor 1980): how can appropriate novel distinctions be generated? Fodor argues that systems cannot increase their logical power (acquire wholly new features) by means of learning: the system's vocabulary and mechanisms must already be able to express the "new" feature, and so that feature has not been "created." This is exemplified by the focus by Schyns et al. on new features being "created" by interpolation between existing values on dimensions: intermediate values are implied by the notion of a dimension and hence are not really "created."[1] However, even though "created" features might thus better be construed as combinations over existing features, there are an infinite number of such combinations – a problem of induction remains.

Despite the fact that this is a critical problem, Schyns et al. fail to address it properly. They state that "categorizations, rather than being based on existing perceptual features, determine the features that enter the representation of objects" (sect. 1.2.4, para. 1). Their position appears circular, because they use feature creation to explain categorisation, but claim that categorisation itself determines feature creation.

Nonetheless, conceptual flexibility needs explication by constraints on acceptable inductions (i.e., combinations of features),[2] and we suggest that this can be achieved by pragmatic factors that guide the appropriate combination of fixed features. Flexibility can be explained via subjects' sensitivity to *relations between* sets of fixed features. Rather than "creating" new features, subjects utilise feature dependencies or correlations, and treat these *as if* they were autonomous features. Sensitivity to such relations is constrained by the purposes of classification and wider cognitive or communicative purposes, and will therefore be dependent on context and task (Franks & Braisby, 1997, submitted).

Hence, the classifications supported by a fixed set of conceptual (or semantic) features can be augmented by pragmatically motivated operations on that set. Whether categorisation exploits semantically or pragmatically derived attributes will therefore depend on the task, context, and the kind of object involved. Natural kind and related concepts, having an elaborate semantic structure, may give rise to less pragmatically guided categorisation than vague or social categories, which have a less determinate semantic structure (see Braisby et al. 1996; Braisby & Franks, 1997, submitted).

Pragmatic factors also explain findings that Schyns et al. take to undermine fixed features. Schyns & Rodet (1997) show that the order in which single feature categories are learned determines whether feature conjunctions are treated as conjunctions or as unitary features. However, different orderings have different pragmatic properties: learning conjunctive categories after simpler ones invites subjects to utilise the importance of attributes from previous categorisations, which is not allowed when conjunctive categories are learned before simpler ones. Thus, categorisation differences need not entail feature creation, because one order may pragmatically "lead" subjects to combine features. A reconstruction of other aspects of the account by Schyns et al. is also suggested by the pragmatic alternative. Their suggestions that task differences and an individual's past history underlie concept flexibility have direct explications in terms of the pragmatics of the act of classification.

In conclusion, Schyns et al. rightly bring to the foreground the importance of flexibility in categorisation, but they err in drawing inappropriate morals for concept representation. As long as existing features can enter into novel combinations, "creationism" remains a myth.

NOTES

1. This simplistic notion does not account for many theoretical features that cannot be thought of as values on dimensions.

2. Such combinatorial operations need not be limited to the simple feature summations that have often been taken to define compositionality (e.g., Hampton 1987). In fact, nonmonotonic operations (e.g., feature negation, modification, and coercion) and operations that permit feature emergence can all be compositional, provided that their outcomes are predictable (see Franks 1995; Frege 1892; Partee 1984).

The development of new functional features by instruction: The case of medical education

Lee R. Brooks

Psychology Department, McMaster University, Hamilton, Ontario, Canada L8S 4K1. **brookslr@mcmaster.ca**
www.science.mcmaster.ca/psychology/faculty.list.html

Abstract: Medical education provides many examples of the development of functional features, but as a response to deliberate instruction. These features require so much specificity and context sensitivity that they seem likely to require the development of new categories of appearances rather than just reweighting old features. A suggested implication is that feature development may help to explain the problematic noticing of features in diagnosis.

The development of new functional features, a theme ably advocated by Schyns, Goldstone & Thibaut, is essential to the development of adult expertise. Medical education provides a particularly compelling example of this process. New students face an array of diagnostic rules that refer to unfamiliar features. Some of these features could initially be regarded as compounds of familiar categories (e.g., butterfly rash). However, all of them, including the apparent compounds, pose a new learning problem. Students are being asked to learn a category of appearances that signal a feature that is functional in making a diagnosis. As an extreme example, in diagnosing lupus erythematosus, a diagnostician is looking for a rash that is butterfly shaped in that it spreads across the nose and fans out on both cheeks, not a rash from applying an allergenic butterfly decal to the face. The rash can be quite variable in appearance, but there are clearly some things it cannot look like. All features named in the rules have a range and form of manifestation that should become familiar and that should be distinguished from appearances so weird as to make the appropriateness of labeling it a functional feature questionable.

This specificity of features has another implication. These functional features become context-specific categories and are not in general interchangeable between diseases, even when the features have the same name. "It has vesicles," applied to a case of contact dermatitis, means the type of vesicles that are characteristic of contact dermatitis, not of herpes simplex. In response to a challenge, dermatologists can immediately explicate the differences and indicate that they were aware of the differences without specifying them. This is analogous to using the same label for two legs on a human and two legs on a bird. Few of us who used the phrase "you have two legs" (in the sense of "get it yourself") would be unaware that the legs in question are human legs, even though that specificity was not mentioned. Anyone who *was* unaware would be in desperate trouble for some of life's other challenges. It is this combination of the concurrent need for specificity and for generalization that makes the generation of functional features a core process that cannot be handled as a simple reweighting of fixed features. This, of course, is Schyns et al.'s main point.

Unlike everyday object categories, the verbal formulations of features in medicine are functional and well practiced. Initial instruction is partly verbal, and extensive verbal interaction continues throughout a medical career. However, as mentioned, the verbal labels are not in any sense exhaustive. A conjecture we could make is that the vocabulary actually used is a compromise between a level of generality that makes communication feasible, such as sentences of less than a thousand words, but sufficiently specific that it points adequately to the referent. In this sense, the technical vocabulary is a signal for the student to begin learning, an authorization for new features as well as the authorization for a new category described by Schyns et al.

Everything to this point illustrates Schyns et al.'s main points. However, it might be worth adding another aspect to their discussion of flexible, learned features. With a process of learning, so many overlapping and context-specific functional features may be developed that under some circumstances, recognizing their occurrence may be problematic. Familiar examples in the context of specific recognition are certain figures used in perceptual research. These figures characteristically have some degradation of information: broken contours (Street figures), heavy shadows (Mooney figures), blurring (Bruner & Potter's gradual bringing into focus), or removal of pixels (Jacoby et al. 1989). (See Kennedy 1993 for an interesting discussion.) People's common experience is that they are initially hard to see, but once seen, they cannot be "unseen." That is, once having had the Dalmatian dog pointed out in a blotchy picture, that interpretation of the same figure emerges effortlessly even when it is presented much later in another context. The fact that it is exactly the same figure (in all of the demonstrations of which I am aware) could mean that seeing an ear or a tail in the figure is so specific that it would be hard to think of these as cases of functional feature extraction.

The same type of phenomenon can be observed in classifying new cases in medicine. Norman et al. (submitted) showed both experts and medical students head and shoulder photographs of patients, each of which showed a key feature that should suggest the correct diagnosis. Three quarters of these pictures were taken from textbook illustrations of the disorder, and all were judged to provide clear examples of the features in question. These experiments provide two lines of evidence that noticing these supposedly obvious features is difficult and is strongly influenced by contextual factors. Both experts and students gained 20% in diagnostic accuracy by having the key, clearly visible features verbally described for them. Both experts and students reported seeing from 15% to 30% more of these features when the correct diagnosis was suggested to them. With the students, we ran controls to show that this facilitation was an increase in sensitivity rather than just response bias. The informal report by experts and students alike was that they had simply not noticed features that seemed clear when they were pointed out. This difficulty of noticing could provide the need for an underlying "coselection" of features and diagnoses much like that hypothesized for the word-superiority effect. In any event, it is hard to understand the missing of "obvious" features if they come from anything like a fixed, high-level feature set. If the fixed features are instead low-level and subsymbolic, then the necessity for a process of high-level feature construction is substantial and cannot be ignored in studies of categorization. This is the dilemma posed for fixed feature theories by Schyns et al.

Fixed versus flexible features in dissociable neural processing subsystems

E. Darcy Burgund and Chad J. Marsolek

Department of Psychology, University of Minnesota, Minneapolis, MN 55455. **burgund@turtle.psych.umn.edu**
levels.psych.umn.edu/lab/index.html

Abstract: Implementational-level evidence of dissociable neural subsystems is a critical element that is missing from the analysis in the Schyns et al. target article. The question of whether fixed or flexible features are used in visual form recognition may have different answers for different subsystems; the evidence that features typically are created during category learning may apply most directly to a specific visual form subsystem.

Schyns, Goldstone & Thibaut argue that higher-level cognitive processes greatly influence the lower-level features that are used during object recognition. Moreover, features may be created during the course of learning new categories. We applaud the computational-level reasoning behind this theory, but we suggest that implementational-level evidence for dissociable neural subsystems is a critical missing element. In particular, the argument that flexible space category learning plays an important role in object recognition may apply to one relevant visual form subsystem of the brain, but not to another.

One important goal in human vision is to recognize *specific categories* of visual form. For example, we can differentially remember the visual forms "BEAR," "bear," and "bear," despite the fact that they belong to the same abstract category of form (e.g., try creating a visual image of each). What sort of feature would be useful for specific category recognition? Presumably, information that is close to the holistic structures of input forms is needed to distinguish specific categories. For example, nearly all the information in "bear" is needed to distinguish it from "bear." Note that such large, holistic features would be useful for distinguishing "BEAR," "bear," and "bear," but would not be useful for marking their commonalities, as needed to achieve a different goal in human vision (see below).

It is highly unlikely that complex, holistic features are fixed; the computational expense of storing all such features for all specific categories that one might learn would be enormous. Also, in line with the reasoning in section 2.3.1 of Schyns et al., it is highly unlikely that very small unstructured primitives (e.g., pixels) are coupled with powerful combination methods to account for the learning of these holistic features. Thus, we conclude that holistic features may be learned and stored as flexible features.

Indeed, recent evidence from a study of visual priming supports this conclusion (Marsolek et al. 1996). Participants first read centrally presented word pairs (one word above the other in each pair) and then completed word stems presented beneath context words in the left or right visual field. Stem-completion priming that was specific to a letter-case match between prime words and test stems was found only when the context word was the same one that had previously appeared above the primed completion word and when the test items were presented directly to the right hemisphere (briefly in the left visual field). Priming that was not letter-case specific did not depend on context or on the hemisphere of direct stimulus presentation. Moreover, this pattern of results was not obtained when the test task was word-stem cued recall, so explicit memory did not account for the interesting results. Two important conclusions may be drawn: first, letter-case specific priming was dependent on right-hemisphere test presentations, thus a subsystem that stores the visually distinctive information needed to differentiate lower- and uppercase versions of the same word may operate more effectively in the right hemisphere than in the left. Second, letter-case specific priming was dependent on visual context; hence this specific visual-form subsystem may not store a word pair as a collection of individual features, but rather, as a single holistic feature. Interestingly, the ability to store holistic features may have been what enabled a specific subsystem to store the novel information (i.e., the new association between two words) effectively.

Another important goal in human vision is to recognize *abstract categories* of form. For example, when reading words for meaning, it is useful to classify different specific inputs, such as "BEAR," "bear," and "bear," as belonging to the same category. Consider the features that would be useful for abstract visual form recognition. Presumably, the information that remains relatively invariant across input forms that belong to the same abstract category should be stored, but other information in the inputs should not. This relatively invariant information is usually present in the parts of any one input form. For example, such information in "BEAR," "bear," and "bear" should include some parts (e.g., a three-pronged vertex halfway up a vertical line on the left side) that are common to those forms but should not include other parts (e.g., a three-pronged vertex in the upper-left) that are not. Indeed, evidence from a study of visual categorization indicates that such relatively invariant information is stored efficiently by an abstract visual form subsystem that operates more effectively in the left hemisphere than in the right (Marsolek 1995). This abstract subsystem may rely on parts-based internal representations that contradict the holistic internal representations needed in a specific visual form subsystem (Marsolek & Burgund, in press).

Unlike specific category recognition, it is unclear whether fixed or flexible features are used to accomplish abstract category recognition. Much of the empirical evidence of feature creation summarized by Schyns et al. appears to be due to processing in a specific visual-form subsystem. For example, in the very clever Martian cells study (sect. 2.2), participants learned either {*xy, x, y*} or {*x, y,*} feature vocabularies, depending on whether they learned *XY* before or after learning *X* and *Y*, respectively. The *xy* feature learned by the group that processed *XY* before *X* and *Y* appears to have been a holistic representation that did not code *x* and *y* as separate and conjoined features, much like the holistic representations that should be useful for specific visual form recognition in one of the two relevant visual form subsystems.

Furthermore, the "alternative" materials that are more likely to give rise to flexible feature creation compared with more "traditional" materials (Table 1), according to Schyns et al., tend to be unfamiliar categories that may be represented effectively without obvious part decompositions. These are just the sort of stimuli that a specific visual form subsystem would be very effective at learning, at least initially. Thus, alternative materials may elicit feature creation more readily than traditional materials because they are more likely to be processed in a specific subsystem. In contrast, stimuli that are readily decomposable into familiar parts may be processed more effectively in an abstract subsystem. If so, it remains unclear whether fixed or flexible features are used to accomplish abstract category recognition.

Hence, the debate over fixed versus flexible features in visual form recognition may apply differently to different subsystems. Implementational evidence of dissociable subsystems may be crucial for a complete resolution to the issues that Schyns et al. raise.

The other hard problem: How to bridge the gap between symbolic and subsymbolic cognition

Axel Cleeremans

Cognitive Science Research Unit, Université Libre de Bruxelles CP 122, 1050 Bruxelles, Belgium; **axcleer@ulb.ac.be**
homepages.ulb.ac.be/~axcleer

Abstract: The constructivist notion that features are purely functional is incompatible with the classical computational metaphor of mind. I suggest that the discontent expressed by Schyns, Goldstone & Thibaut about fixed-features theories of categorization reflects the growing impact of connectionism. Their perspective is similar to recent research on implicit learning, consciousness, and development. A hard problem remains, however: how to bridge the gap between subsymbolic and symbolic cognition.

Schyns, Goldstone & Thibaut's target article touches on deep and fundamental issues in cognitive science, and it is in this way that their claims are relevant to several domains beyond categorization, including consciousness (Cleeremans, in press; Perruchet, in press), development (Clark & Karmiloff-Smith 1993), and implicit learning – the process by which we can learn without intending to do so and without awareness of the resulting knowledge. In this commentary, I argue that the constructivist view of cognition advanced by Schyns et al. is consistent with new perspectives on implicit learning. The common theme is that the classical framework is under attack in both the categorization and implicit learning fields. The classical framework is grounded in the idea that cognition is about manipulating (symbolic) representations (Newell & Simon 1972). It is at the core of many models of performance in a wide variety of domains, and also constitutes the underlying theoretical framework of most experimental work in cognitive science.

The key classical notion is that cognition involves "manipulating" representations. This assumption entails that (1) at least some primitive representations preexist (before the onset of a learning or a categorization episode, or before the onset of development); (2) that these representations are essentially static and fixed; and (3) that they exist independently of the processor that manipulates them. This makes it very natural to imagine that features, as representations that have diagnostic value, somehow exist independent of their use in a given context. In contrast, Schyns et al. propose that features do not exist before they are needed and get created. Features thus have a purely functional existence determined by their relevance in a specific task context as well as by the learner's previous experience.

The same dissatisfaction with the traditional metaphor of mind is now emerging in implicit learning research. In contrast to typical categorization tasks, implicit learning tasks usually involve precisely the kind of complex stimuli that Schyns et al. wish to see used more often in their own domain. Beyond stimulus complexity, however, the most important difference is that participants in implicit learning tasks are never explicitly instructed to encode the regularities contained in the material. This, however, does not prevent them from becoming sensitive to this information, although only in an indirect, implicit way. For example, people exposed to sequentially structured sequences of stimuli presented in the context of a choice reaction time task become progressively more sensitive to the rules used to generate the sequences (as suggested by their faster responses to structured as opposed to unstructured material) despite the fact that (1) the task does not require them to develop this sensitivity, (2) the instructions do not mention the existence of a rule system, and (3) their performance in comparable direct tests of their knowledge of the rule system remains very limited (Jiménez et al. 1996).

Similar patterns of results have been repeatedly obtained with other implicit learning tasks, such as artificial grammar learning (Reber 1993) or system control (Berry & Dienes 1993). In contrast to typical concept learning and categorization tasks, implicit learning tasks never require participants to develop fully explicit symbolic representations of the relevant information. Participants must acquire "features," however, because their performance indicates successful (but indirect) discrimination between different categories of stimuli. The fact that they can do so without having developed the corresponding explicit, verbalizable, symbolic knowledge that is made mandatory by the demand characteristics of categorization tasks suggests that implicit learning tasks can probably be thought of as involving the initial stages of categorization, that is, precisely the kind of feature creation processes described as centrally important by Schyns et al.

I have argued elsewhere (Cleeremans 1994a; in press) that the connectionist framework probably provides us with the best metaphor for thinking about how such implicit knowledge can come to influence performance without having the properties of symbolic representation. Connectionist models indeed excel at representing complex stimuli as graded, continuous patterns that can preserve much of the variability contained in the input while at the same time providing enough structure to satisfy the demands of the task (Cleeremans 1993; 1994b). Furthermore, connectionist models make it clear how such knowledge can develop based on subsymbolic processing and representational principles.

To summarize, the target article makes it clear that theories of categorization based on the traditional computational metaphor of mind (roughly, processes that manipulate representations) are fundamentally flawed. A growing dissatisfaction with the computational metaphor is also emerging in consciousness and implicit learning research, and for essentially the same reasons: the relative inflexibility of symbolic representation, the artificial character of a nevertheless deeply ingrained distinction between processes and representations, and the weaknesses of classical systems in accounting for change and for the emergence of new forms. There is little doubt that connectionism and other similar dynamic (Port & van Gelder 1995) approaches offer many fundamental insights into how one may start to address these limitations of classical symbolic systems. The problem with Schyns et al.'s approach, however, is that they like many others fail to offer any specific discussion of "the other hard problem" that is common to the categorization, implicit learning, consciousness, and development fields: how to bridge the gap that separates subsymbolic and symbolic cognition. In this respect, representational redescription, as proposed by Clark and Karmiloff-Smith (1993), is an interesting suggestion in that it leaves open the possibility that multiple recodings (possibly cast at different levels of complexity) of the same input remain available for further processing, a property that is undoubtedly central to flexibility in any learning system.

ACKNOWLEDGMENT
The author is a Research Associate of the National Fund for Scientific Research (Belgium).

Flexible feature creation: Child's play?

Gedeon Deák

Department of Psychology and Human Development, Vanderbilt University, Nashville, TN 37203. **deakgo@ctrvax.vanderbilt.edu**
peabody.vanderbilt.edu/psych_and_hd/faculty/deakg

Abstract: Schyns, Goldstone & Thibaut's argument is evaluated from a developmental perspective. Theoretically, feature creation is not necessarily problematic; this view derives from the assumption of innate content (primitive feature sets). Alternative assumptions (e.g., Piaget's theory) are possible. Preschool children readily search for novel features in response to task demands. This is compatible with functionalist approaches, but not the rationalist ones criticized by the authors.

What is innate? Alternative views. Schyns, Goldstone & Thibaut are to be commended for attempting to address a critical problem: if the symbol-grounding problem is solved by primitive feature sets, how can we account for (sporadic) flexible and creative induction based on nonprimitive features? If primitives are used productively, what are the production rules? More radically, when and how do people create new features? The difficulty of the problem is underscored by Fodor's (1975) reductio that all complex concepts are innate. Although the Schyns's et al. approach is more sober, primitive constituents are regarded as a necessary evil (sect. 1.1, para. 4). Are primitive vocabularies necessary? Can the problem be solved without innate content (see Braine 1994)?

The symbol grounding problem is inherently developmental, and developmental analysis yields an alternative formulation: Piaget was aware of the problem, and did not believe in primitive content (i.e., features or concepts) but rather in primitive action patterns. Specifically, a set of innate reflexes evolve into controlled action schemes (later internalized as mental schemes). By analogy to Gibson's differentiation hypothesis (Gibson & Gibson 1955), the infant's ever-finer differentiation of the physical world is reflected in differentiated action responses. Primitives are not content, but structure and *process* – specifically, perceptual learning and motor learning routines. My point is not to advocate Piaget's theory, but to remind us that what is innate might not be primitive feature sets. Whether or not primitive features are assumed, however, feature creation remains to be specified.

The growth of flexibility. Schyns et al. link feature creation to flexibility, a welcome observation. Flexibility is most clearly construed from a functionalist position, by questioning how subjects elect specific features appropriate for different tasks. Subjects might select either previously conceived aspects or novel aspects of the stimulus array. If the latter is typically more effortful and uncertain than the former, there will be a trade-off: feature search will occur only when existing features are ill-suited to a problem. Schyns et al. consider how stimulus characteristics affect feature creation, but they do not address how task and context facilitate or inhibit feature creation. For example, Bransford et al. (1989) suggest that conceptual highlighting of contrasting features might promote feature creation. Deák and Bauer (1996) found that preschoolers search for subtle (and presumably novel) features in certain task contexts, given sufficiently complex stimulus items (this qualification is consistent with Schyns et al.'s argument about stimulus characteristics; see sect. 2.5).

In terms of development, Schyns et al. imply that apprehension of novel features poses particular difficulties for young children (sect. 3.3.2, para. 3–4). However, learning novel feature contrasts is so critical for young children, it would be surprising if they were deficient in it. Schyns et al. seem to know this, and conclude correctly that current popular approaches (i.e., innate theories; perceptual biases) are not enough to specify how children create or select relevant features from an array. Theoretical shortcomings aside (see Deák 1995; in preparation), both popular approaches are empirically wanting. Consider the view that children's feature selection is governed by innate perceptual biases: Deák (1995) qualifies or disconfirms the count noun/shape bias proposed by Landau et al. (1988). A comprehensive review of the literature reveals that preschool children are not *generally* biased to weigh some features over others. Rather, they select stimulus features associated with a particular induction problem (Deák 1995). That is, preschool children are genuinely flexible in shifting attention to features (or combination of features) as task demands change. Moreover, it appears that feature selection and feature creation are closely related, and well-established, by age 3. I will briefly describe data (which inform Schyns et al.'s general position) consistent with this argument.

Deák (1995) found that 3- to 6-year-olds successively attend to different combinations of features in response to different induction tasks: when told that an unfamiliar, complex object "has a toggle," 3- to 6-year-olds extend that fact to another object with the same unfamiliar part. When told that the former object is "made of mylar," children extend that fact to a different object made of the same (unfamiliar) material. How is this relevant to the position stated by Schyns et al., besides being inconsistent with the perceptual bias approach? First, in response to ". . . has a . . ." facts, children apparently search for an unfamiliar part of the object, then seek another object with the same part. Parts were novel, and the objects were complex, with many varying features. Thus, children apprehend and reason about novel features in spite of irrelevant varying information (contrary to Schyns et al., sect. 3.3.2). Second, responses to ". . . made of . . ." facts demonstrate that preschoolers reliably make generalizations about novel kinds of material. This is striking partly because 3-year-olds are believed to lack a coherent concept of material kind (Dickinson 1989). Thus, children "created" a new material feature, although by conventional accounts no feature space existed to be subdivided! This illustrates how task constraints might permit young children to induce novel features or feature combinations. Clues to the nature of the task – for example, the phrases "has a" and "made of" – effectively limit the hypothesis space preschoolers consider.

In sum, assuming that "what is innate" is either nascent theories or rigid perceptual biases (i.e., innate content *qua* primitive features) provides no account of either flexible feature selection or feature creation. A functionalist framework, in contrast, assumes that children have procedures for matching (or learning to match) known or novel features with exigencies of the task at hand. The emergence and nature of these procedures remain to be understood.

Flexible categorization requires the creation of relational features

Peter F. Dominey

Vision et Motricite, Inserm U94 and Institut des Sciences Cognitives CNRS UPR 9075, 69373 Lyon, France. **dominey@lyon151.inserm.fr**

Abstract: Flexible categorization clearly requires an adaptive component, but at what level of representation? We have investigated categorization in sequence learning that requires the extraction of abstract rules, but no modification of sensory primitives. This motivates the need to make explicit the distinction between sensory-level "atomic" features as opposed to concept-level "abstract" features, and the proposal that flexible categorization probably relies on learning at the abstract feature level.

It is clear that a fixed and finite feature set cannot anticipate all possible categorizations. Schyns et al. thus set out to establish a framework in which the feature set can be augmented with new features. Part of the burden of proof on the authors is to demonstrate a case in which new feature creation can clearly be dissociated from the weighting or combination of existing features (sect. 2.6). In this regard, we study a form of categorization of dynamic objects – sequences – in which the defining features for categorization are configurational relations between elements or features, independent of specific features themselves. I develop this point as an example of a case in which the creation of novel features cannot be a simple combination of existing features. I note also that in this and the subsequent points, the dynamic object results generalize to static objects.

We have recently described a dissociation between cognitive processes for learning surface structure and abstract structure of sensorimotor sequences. Surface structure is simply the serial order of the elements in the sequence, whereas abstract structure is defined in terms of positional relations between elements that repeat in a sequence. Thus, the two sequences ABCBAC and DEFEDF share the same abstract structure (123213) but have different surface structures, and are thus isomorphic. We have demonstrated in a serial reaction time task that although surface structure can be learned under implicit conditions, abstract structure can only be learned under explicit conditions (Dominey et al. 1995; 1997; in press). A hallmark of abstract structure learning is the capacity to transfer knowledge of the abstract structure to new, isomorphic sequences. Specifically, when subjects trained with sequences such as ABCBAC are exposed to a new isomorphic sequence DEFEDF, they can transfer knowledge of the shared abstract structure 123213. This transfer yields significant performance benefits for the elements in the new isomorphic sequence that are predictable by the abstract structure. Thus, by means of their training, these subjects have gained category knowledge that allows them to "categorize" sequences as either belonging to the isomorphic set or not. We thus have a condition in which previous experience significantly modifies the immediate appearance of dynamic objects or sequences.

The question remains: Has a new feature been created? A response can be provided from simulation studies of abstract structure learning. In these studies, a neural network model is capable of learning and transferring abstract structure between isomorphic sequences (Dominey 1995). In the model, the set of sensory-perceptual inputs or atomic features was fixed. Flexibility came from an adaptive capacity to represent arbitrary relations between atomic features. This type of learning allows it to be recognized that ABCBAC and DEFEDF share a common abstract structure of internal repetition.

In saying that features are both elementary stimulus properties yet little different from concepts, Schyns et al. may have blurred a useful distinction between levels of representation. I thus suggest a slight nuance: "Atomic" features correspond to low-level (sensory) primitives, and "relational" or "abstract" features are defined in terms of configurational relations between atomic features, as in the abstract structure learning already described. In this framework, relational feature creation does not involve the extraction of a new feature dimension explicitly represented in the stimulus at the atomic or perceptual level, but instead entails the extraction of a relation between such features, independent of the feature instances themselves. Such a capability is clearly a key element in a flexible categorization scheme.

Flexible features, connectionism, and computational learning theory

Georg Dorffner

Department of Medical Cybernetics and Artificial Intelligence, University of Vienna, A-1010 Vienna, Austria.
georg@ai.univie.ac.at www.ai.univie.ac.at/oefai/nn/georg.html

Abstract: This commentary is an elaboration on Schyns, Goldstone & Thibaut's proposal for flexible features in categorization in the light of three areas not explicitly discussed by the authors: connectionist models of categorization, computational learning theory, and constructivist theories of the mind. In general, the authors' proposal is strongly supported, paving the way for model extensions and for interesting novel cognitive research. Nor is the authors' proposal incompatible with theories positing some fixed set of features.

Schyns, Goldstone & Thibaut's proposal is remarkable and important in many ways. It is remarkable because it touches on issues that have come up in other domains of cognitive science but have received surprisingly little attention in the literature about categorization. It is thus worth investigating Schyns et al.'s view on features in the context of two areas they have not explicitly discussed: connectionist models of categorization and the theory of learnability.

Connectionist models of categorization can be divided into at least two strands: models based on backpropagation in MLPs (multilayer perceptrons) (Harnad 1987; Rumelhart et al. 1986), and models based on more localized responses in radial-basis function networks (Kruschke 1991) and variations of competitive learning (Dorffner et al. 1996; Grossberg & Stone 1986). MLPs are a variant of neural networks that apply a weighting scheme to features, whereas the other kinds of model compute a distance measure between a prototype (the weight vector) and an input pattern.

With respect to the former, it is interesting to note that hidden units in multilayer perceptrons can be interpreted in exactly the way as Schyns et al. suggest. To arrive at a complex categorization, hidden units develop higher-order combinations of input features that are influenced by the categorization task itself. Put differently, hidden units can be regarded as higher-level features on which subsequent categorization is based (compare Bishop 1995, pp. 226ff). These fulfill Schyns et al.'s requirements, especially the fact that they develop with the help of feedback from the categorization task.

Models using localized responses appear to be more limited in light of the proposal by Schyns et al. Existing models indeed work on a fixed set of input features and apply one level of weighting through the distance measure. Thus, Schyns et al. show the need for a substantial extension. One way to approach hierarchical categorization is to add one or several levels of categories between the input and the final category. In other words, categories cannot be the result of only weighting input features; more complex features themselves can be generated by a categorization process. We have recently argued along similar lines, even though our model is not fully implemented (Dorffner 1997). Our approach agrees with Schyns et al. that "there is little difference between concepts and features" (sect. 4, para. 1). Important in such an approach would be top-down feedback from higher-level categorization to feature-level categorization. This could be implemented in the framework of the adaptive resonance theory of Grossberg (1976) through the recuritment of new feature categories.

Computational learning theory (Valiant 1984) can shed more light on Schyns et al. theory. It is safe to assume that for categorization there must exist a level where features are fixed, the level of peripheral sensory features (e.g., retinal activations). According to computational learning theory, given n such fixed features, if a learner is able to represent all 2^{2^n} dichotomies (assuming that features are binary and the system is learning to distinguish only two categories) then learning is impossible. Thus, substantial bias is necessary to constrain learning, as Schyns et al. acknowledge in their discussion (sect. 3.3). They suggest that the top-down feedback during categorization can provide such a bias in constraining the features that are constructed. This view has interesting implications in suggesting that the environment providing the feedback with respect to categories itself supplies considerable bias. In other words, a large proportion of the 2^{2^n} theoretically possible dichotomies are ruled out by the fact that the environment does not provide feedback for them to be useful. Environmental constraints are also suggested by the work of Elman et al. (1996) on innateness.

Such bias still seems too small to make the learning problem tractable, so additional constraints on possible features must be innate. In this sense, Schyns et al.'s theory is compatible with most of the literature positing fixed feature sets (e.g., the papers discussed in sect. 2.3). If learning (in this case, categorization) is to be tractable, a mixture of (a) fixed, (b) preconditioned but adaptive, and (c) fully adaptive features are needed. Future research will have to determine the relative proportions.

The proposal of Schyns et al. is an important and necessary amendment to many cognitive models and theories, pointing to important extensions on the one hand and paving the way for fascinating new research on the other.

Things are what they seem

Shimon Edelman

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

Abstract: The learnability of features and their dependence on task and context do not rule out the possibility that primitives used for constructing new features are as small as pixels, nor that they are as large as object parts, or even entire objects. In fact, the simplest approach to feature acquisition may be to treat objects not as if they are composed of unknown primitives according to unknown rules, but rather as if they are what they seem: patterns of atomic features, standing in various similarity relationships to other objects, which serve as holistic features.

I sympathize with the notion of feature learnability, for which Schyns, Goldstone & Thibaut marshal unimpeachable arguments, and which they support by an impressive range of experimental evidence. I would like to take issue, however, with the analysis of the possible approaches to the *implementation* of learnable features, offered by Schyns et al. in section 2.3. That section contains three statements that affect crucially the theoretical orientation of the rest of the article; I discuss each of these in turn, then suggest an alternative view of flexible features, along with some practical conclusions.

Putative computational limitations of unstructured primitives. In section 2.4.1, Schyns et al. claim that "it is not practically feasible (although it is logically possible) to extract relevant categorization features from pixel-based (or similarly unstructured) representations of the input." Most approaches to dimensionality reduction (discussed later, in section 3.4.2) indeed cannot handle the extraction of nonlinear or otherwise complicated category structure from raw images. At least one recent method, however, does show promise in this respect, by combining a nonlinear trainable mechanism for function approximation (such as a multilayer perceptron network) with the imposition of category-related constraints during learning (Intrator & Edelman 1997). According to this method, the target manifold (the subspace that contains the data points and is embedded in the multidimensional space of raw features) is captured by teaching the network (1) to represent between-categories variation, corresponding to the directions tangent to the manifold, and (2) to ignore within-category variation, corresponding to the directions orthogonal to the manifold. This makes it possible to extract a topologically faithful replica of a two-dimensional nonlinear subspace occupied by a class of objects (human heads), from a feature space of a thousand or so dimensions.

Putative representational limitations of structured primitives. In section 2.4.2, Schyns et al. state that "any large scale, highly structured set of primitives is bound to be too coarse to detect (and internally represent) all of the distinctions that might be required by different categories of objects." To support this statement, Schyns et al. give the example of the purported "clumsiness" of a set of bar-like features in representing curved-line shapes such as ellipses. Our intuition, however, is a poor guide in this matter, and is misled by an appeal to the notion of clumsiness, which serves here as what D. C. Dennett calls an "intuition pump." To dispel the false intuitions, one may observe (Edelman 1995) that overlapping graded-profile "feature detectors" similar to the receptive fields found in early vision can support directly the categorization of seemingly unlikely stimuli. For example, *circularly symmetric* Gaussian receptive fields can support discrimination of Verniers formed by *straight* lines, at a hyperacuity level (Poggio et al. 1992). Analogous mechanism may be at work at a much higher level: as shown in Edelman and Duvdevani-Bar (1997), everyday objects can be represented (and recognized and categorized) by their holistic similarities[1] to a number of reference shapes of comparable structural complexity.

Feature extraction as a metatheoretic enterprise. My third point of contention is Schyns et al.'s view that "the task of the subject creating new functional features for categorization is not substantially different from the task of the scientist creating a *componential* theory of recognition" (sect. 2.4.3, emphasis added). The problem here lies in the tacit assumption that feature extraction must be patterned on the reductionist notion of the need to take something apart to be able to understand it. To avoid a lengthy digression into the philosophy of reductionism in science, I merely point out that in certain situations the only understanding that seems to be available is cast in terms of a very complex, frequently probabilistic, mapping between observables. Likewise, in the dimensionality reduction example cited above, the features turn out to be complex patterns of weights gleaned from a network following its training; there is nothing to guarantee the possibility of making sense of such features in image-wise "componential" terms.

Summary. There is a way to maintain a flexible attitude toward features that does not involve the need to create a "componential theory" of recognition for every given task. First, recent results in visual modeling suggest that it is possible to learn the appearances of entire objects and to use the resulting memory traces subsequently as holistic features. Second, advanced methods for dimensionality reduction can extract useful (yet in a traditional mereological sense unanalyzed) features directly from image data. These can augment appearance-based holistic features, which uphold the notion that, in terms of representation, things are by and large what they seem – a possibility that is rather economical, albeit not too appealing to a radical reductionist.

ACKNOWLEDGMENT

I thank Nathan Intrator for many stimulating discussions of various issues having to do with features and representations. For additional material, including preprints of relevant papers, see www.ai.mit.edu/~edelman.

NOTE

1. That is, similarities computed over entire "unanalyzed" objects, and not over their descriptions in terms of, for example, spatially distinct parts.

New-feature learning: How common is it?

Robert M. French[a] and Mark Weaver[b]

[a]*Psychology Department, University of Liege, 4000 Liege, Belgium;* [b]*Corvus Development, Inc., Ann Arbor, MI.* **rfrench@ulg.ac.be**
weaver@corvusdev.com

Abstract: The fixed-feature viewpoint Schyns et al. are opposing is not a widely held theoretical position but rather a working assumption of cognitive psychologists – and thus a straw man. We accept their demonstration of new-feature acquisition, but question its ubiquity in category learning. We suggest that new-feature learning (at least in adults) is rarer and more difficult than the authors suggest.

Schyns et al. are in the main justified in their skepticism about conventional approaches to category learning – to work with conveniently small numbers of fixed, high-level features is to make a set of simplifying assumptions that have the potential of missing much of the complexity of the category learning process. However, we find that they oppose a straw man and that their conclusions go beyond the supporting evidence. Despite this, the work Schyns et al. present has important implications that they themselves have only begun to explore, and it provides valuable insights into the learning of new categories through the acquisition of the feature sets.

To begin where the target article does, Schyns et al. contrast their position to a fixed-feature alternative they describe as "an influential approach." The problem with this is that they are not clear about the difference between fixed features as a theoretical position and fixed features as a working assumption. As a theoretical position, a "fixed-feature stance" is clearly a straw man. Features must, of course, come from somewhere, and because it is surpassingly unlikely that we are born with all the features we will ever need (like the eggs in baby girls' ovaries), one must presume that features are, in fact, learned. However, there is no evidence that the "fixed-feature stance" has been seriously put forward as a theoretical position. It is, rather, a (widespread) working assumption. Is this assumption valid? Schyns et al. argue that it is not. However, even if we accept their claim that they have demonstrated the acquisition of new features during category learning – which we do – it does not follow that the acquisition of new features is a *typical* part of category learning. In fact, it is reasonable to expect that, just as adult speakers of English have in place all the phonemes they will ever need to represent English words (even the many thousands outside their current vocabularies), they also have in place all the primitive features needed for learning new categories in many domains. What is more, the exceptional character of the stimuli in the authors' "Martian cells" and "Martian rocks" experiments is a tacit acknowledgment of how hard it is to force subjects outside their existing areas of "featural competence." Thus, we argue that the widespread assumption of fixed features in category learning experiments *is* supported in the general case of adult category learning.

One of the most promising but elusive points in the target article involves the distinction between "primitive" and high-level, "functional" features. On the one hand, the authors point out that almost *any* feature is composed of lower-level features – at least until one gets all the way down, for example, to the edge detectors in low-level visual cortex. This makes the distinction between "primitive" and "functional" features difficult, if not impossible, to maintain. On the other hand, the authors hand much of their argument on the distinction: "If a primitive set of features can capture all the regularities and categorizations accommodated by the functional features . . . then the hypothesis that feature creation is needed to allow a system to represent object properties it was previously incapable of cannot be maintained."

One way out of this apparent contradiction is to think of both "primitive" and "structured" features as hierarchically organized building blocks from which new categories and concepts can be constructed. As we have argued, because many categories will be able to share the same general-purpose building blocks (e.g., something like Biederman's (1987) geons), category learning can often proceed by assembling existing building blocks. In other cases (as with the "Martian rocks" and "Martian cells," we suspect), new building blocks are needed, but they are few in number and may hence be acquired as a side effect of category learning without significantly impeding the process. At the other extreme – something Schyns et al. have not acknowledged – are situations where appropriate features are generally missing. In addition to the case of infants confronting a bewildering world, one might add those cases in which adult learners attempt to master new and unfamiliar domains of knowledge. Examples of this kind of feature learning might include the experience of a native English speaker learning to hear the tonal phonemes of Mandarin Chinese, or a neophyte mushroom gatherer learning to discriminate edible and poisonous varieties). Feature (and, therefore, category) learning in such cases is often slow, difficult, and even painful.

Finally, we suggest that, ironically, Schyns et al. may have focused their demonstration of new-feature learning precisely where it is most difficult to observe – namely, with readily perceived, concrete objects. Both evolutionary constraints (favoring concrete categories) and universally shared experience with objects in the world suggest that by the time of adulthood, most of the primitives that humans need to handle concrete categories will already have been learned – which is why we argue that a fixed-feature approach for the learning of concrete categories is a reasonable working assumption. On the other hand, as categories grow more abstract (as in chess playing, mathematics, classical music, architecture, aeronautics, etc.) – and therefore less universal and directly related to needs of survival in the physical world – one might expect a far greater degree of new-feature creation. Thus, the slow building of expertise in abstract domains may present a better opportunity for studying the interaction between feature acquisition and category learning.

In conclusion, we feel that Schyns et al. have needlessly presented their argument in contrast to a straw man opponent. Although the widespread working assumption of fixed features in category learning *is* justified, we agree that the acquisition of new features is an important topic that has to date received scant attention in both empirical research and category learning theory. The target article represents a useful start in both of these areas.

ACKNOWLEDGMENT
This work was supported in part by Belgian PAI grant no. P4/19.

Building block dilemmas

Christopher Gauker

Department of Philosophy, University of Cincinnati, Cincinnati, OH 45221-0374. **gaukercp@email.uc.edu**
ucaswww.mcm.edu/philosophy/gauker/

Abstract: Feature-based theories of concept formation face two dilemmas. First, for many natural concepts, it is hard to see how the concepts of the features could be developmentally more basic. Second, concept formation must be guided by "abstraction heuristics," but these can be neither universal principles of rational thought nor natural conventions.

When we ask how we can tell that something is a bird by just looking, it is natural to suppose that we do it by recognizing the features of birds that tend to distinguish them from other kinds of things. This commonsense theory of object recognition suggests a theory of concept formation: children form their general concepts of observable types by learning which observable features of things tend to characterize the types. If this is right, then surely the concepts of the features must be developmentally more basic than the concepts whose genesis we aim to explain. Maybe they are even innate.

Any such theory of conceptual development faces an obvious dilemma. If we confine ourselves to concepts that might be developmentally more basic, we might list such concepts as *small, mobile,* and *blue* – maybe even *causes* – maybe even *endures unseen.* However, it is hard to see how birds might be distinguished from other kinds of things only by means of concepts such as these. If, on the other hand, we list the features that tend to characterize specifically birds, we might list such features as flying, being feathered, and resting on tree branches, but it is doubtful whether the concept *feathers,* the concept *flight,* and so on, are developmentally more basic than the concept *bird.* No doubt children can in some sense *see* flight and feathers before they possess the general concept *bird,* but we cannot explain the formation of general concepts in terms of sensory contact alone without appealing to some kind of mental act by which that which is perceived is grasped.

One response to this dilemma is to grab the first horn and try to identify a repertoire of developmentally basic concepts adequate for drawing the necessary distinctions. Biederman's geon theory (1987), for example, can be viewed as such a response. The other response is to grab the second horn and argue that the concepts of the features need not be developmentally more basic. This is how we should view the proposal of Schyns, Goldstone & Thibaut. The trouble is that Schyns et al. do not explain how novel feature concepts might be formed. They explain that new features are created when a categorization task requires them (sect. 2.4) and that new ones may be required because categorization is more efficient when only a small number of features are relevant (sect. 3.4.1), but they have nothing to say about how truly new feature concepts may be created. Dimensionality reduction (sect. 3.4.2) is no answer, because that is just a way of defining new features in terms of old. This is ground for criticism bad, because if we had an explanation of the genesis of feature concepts, then we might find that we could explain in the same way the genesis of those concepts whose genesis was supposed to be explained in terms of feature concepts. That seems likely given that "there is little principled distinction" between feature concepts and other concepts (sect. 2.8, last para.).

Even if this dilemma can be evaded, a further dilemma arises as well. As Schyns et al. acknowledge (sect. 1.2.3), one thing a theory of conceptual development has to explain is how children learn to apply category words to roughly the same sorts of things as others who have learned the language. After limited exposure to uses of the word "chair," children will eventually reach a point where they will recognize, without being told, that a certain three-legged chair is a "chair," that an ottoman is not, and that a bar stool is a difficult case. No two people abstract their general concept *chair* from exactly the same class of instances, and yet people tend to agree in their applications of the word "chair." If concepts are indeed the product of abstraction, then the explanation has to be that there are certain shared principles of abstraction. These may be general rules that are applied in the course of abstraction, or they may be conceivable only as aspects of the mind's innate physical architecture. Schyns et al. refer to "biases" (sect. 3.3), but a more general term is desirable. I call them *abstraction heuristics.*

The further dilemma concerns the status of these abstraction heuristics as principles of thought. One possibility is that they are universal principles of rational thought, shared by every possible thinking thing to the extent that it is rational. Against this, one can imagine systems of concepts that seem, by our lights, unnatural and gerrymandered, but in which the elements fit together in a way that supports inductive and other sorts of reasoning. The other possibility is that abstraction heuristics are merely natural conventions, like mating rituals in birds. Human beings might share them only because otherwise we could not learn to coordinate our activities by means of language. Because Martians would likewise learn their languages only because they shared abstraction heuristics with one another, we could not understand their language in the same way if we did not happen to share their abstraction heuristics. Not only could we never learn to speak the Martian language, which is not surprising, but arguably we could never even have a good theory of how the Martians managed to cooperate with one another by means of it. The doubt about this is whether we really want to say that facts, such as how a team of Martians builds a rocket, are necessarily invisible to us.

The conclusion I draw from these dilemmas is that it is a mistake to think of concept formation as a process of assembling a structure of feature concepts. A better strategy, in my opinion, would be to recognize a rich capacity for nonconceptual thought, and then to explain language learning as a product of this nonconceptual thought, and finally equate concept formation with mastering a word (see Gauker 1994).

Self-organizing features and categories through attentive resonance

Stephen Grossberg

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

Abstract: Because "people *create* features to subserve the representation and categorization of objects" (abstract) Schyns et al. "provide an account of feature learning in which the components of a representation have close ties to the categorization history of the organism" (sect. 1.1). This commentary surveys self-organizing neural models that clarify this process. These models suggest how "top-down information should constrain the search for relevant dimensions/features of categorization" (sect. 3.4.2).

Adaptive resonance theory (ART) models illustrate some of the themes of Schyns, Goldstone & Thibaut. An ART model of visual object recognition takes its inputs from prestriate boundary groupings and surface representations, and categorizes them in temporal and prefrontal cortices (Carpenter & Grossberg 1991; 1993; Grossberg 1994; Grossberg & Merrill 1996; Grossberg et al. 1994). When a supervised ART, or ARTMAP, model responds, for instance, to a boundary grouping, it learns to bind together features of the grouping via a bottom-up adaptive filter that activates a category representing this new combination of features. This bundle represents a new "emergent feature," or prototype, that is characteristic of the category.

The active category sends top-down signals back toward the boundary grouping. These top-down signals also encode the prototype, which is matched against whatever boundary grouping is present, and thereby generates a focus of attention. If the match is good enough, the system learns to incorporate a novel boundary grouping into the learned category prototype. Such learning implicitly incorporates all the information ever experienced by the learning subject, because the category that is chosen depends on all the available prototypes, and the change in each prototype depends on all groupings previously experienced by that category. In this way, a new "emergent feature" is learned within each category, and this prototype dynamically reorganizes cell responses, through top-down attentional focusing, in an experience- and context-dependent way. New "features" can be learned in this way, and can influence the perceptual stages that create color afterimages and figure/ground segregation (Francis & Grossberg 1996; Grossberg 1994).

Such features can be learned to classify textures or textured scenes to which the system is exposed (Grossberg & Williamson 1996) by discovering predictive combinations of either boundary and surface properties for classification, or fuzzy rules (Carpenter & Grossberg 1991) such as "any height between 70 and 80 cm" (sect. 3.2). This interplay between bottom-up and top-down learning and attention emphasizes the "intrinsic futility of searching for the boundary between perception and conception" (sect. 2.5).

Superimposed on this process is another level of categorization that binds multiple categories into a final prediction, much as multiple parts contribute to an object whole, or multiple visual fonts predict the verbal name of a letter. Thus, the lower-level categories can be viewed as new features that contribute to the larger category. This higher category level can selectively bind together, or fuse, certain combinations of new features in one context (e.g., spatial frequencies or boundary/surface combinations), and different combinations in others (Asfour et al. 1993).

Such learning attempts to generate the largest categories consistent with environmental feedback, thereby conserving memory resources, much the way stimulus dimensions that are originally processed together by children are later differentiated. A process called vigilance control dynamically alters the system's sensitivity to environmental features based on its predictive success in increasingly complex environments (Carpenter & Grossberg 1991; 1993). When fast learning is allowed, "different histories of categorization generate different feature spaces to encode similarities and contrasts between objects" (sect. 2.2, para. 4) even though all the feature spaces tend to generate similar recognition accuracy if the environment is sufficiently broadly sampled (Carpenter & Grossberg 1991).

ART models do not require predetermined "geons." Biederman (1987) invoked geons to explain data from studies that delete line segments from objects. They can be explained instead in terms of amodal completion of missing boundary segments before they are categorized (Grossberg 1987, sect. 20). ART also allows the self-organized learning of invariant three-dimensional object categories from two-dimensional-view categories (Bradski & Grossberg 1995), as in the multiple-views approach, using a "boundary-based scheme" (Biederman 1987, p. 11) that avoids elastic three-dimensional templates by preprocessing emergent boundary groupings using invariance filtering and optimal coarse coding before they are categorized.

New features can also emerge through preattentive perceptual learning, by using the adaptive horizontal interactions and bottom-up and top-down adaptive filters that occur as early as LGN (lateral geniculate nucleus) and cortical area V1 (Grossberg 1995; Grossberg et al. 1997). ART models suggest that similar types of top-down learning and attention regulate the emergence of new features on multiple levels of thalamocortical processing, from specific thalamic nuclei, like LGN, to prefrontal cortex.

Real-world categories don't allow uniform feature spaces – not just across categories but within categories also

Ulrike Hahn and Nick Chater

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

Abstract: The Schyns et al. target article demonstrates that different classifications entail different representations, implying "flexible space learning." We argue that flexibility is required even at the within-category level.

We welcome this timely emphasis on the need for "flexible spaces" in categorization. In this commentary, we ask how far the points made must be taken. The target article stresses that new features are required by new categorizations and that assuming a single, fixed, object representation suited for all possible classifications is unrealistic. Our work on similarity-based categorization (Hahn 1996; Hahn & Chater, in press) has stressed that fixed, uniform representations are inappropriate even *within* a category. Where Schyns et al. emphasize that, for example, a single fixed-length vector for object representation is insufficient across categories, our aim has been to show how a single fixed-length vector representation is overly restrictive even for a single category. That "uniform feature spaces" are insufficient even within a category becomes apparent with the analysis of real-world materials such as legal cases (Hahn 1996), but more real-world categories suffice equally: imagine encountering a particular chair, one with a back rest and four legs; the next exemplar encountered might have armrests too – a new "dimension" that comes into play only at this point, yet another chair might have a swivel base instead of four legs, and so on. The "feature space" for the category emerges only gradually as more and more examples are encountered. The crucial point, however, is that for many categories, if not most, it is never definitively fixed. New, previously unanticipated variations can arise all the time. The problem is not simply that of encountering a sufficient number of exemplars to allow determination of the space of possibilities, because this space generally is not bounded (at least from the agent perspective). This follows from considering a key difficulty for rule-based systems, that rules – whether attempting to govern everyday, commonsense knowledge or specialist domains such as law – almost always admit of exceptions (Hahn & Vogel 1997; Oaksford & Chater 1991; Reiter 1980). These exceptions, which are both unforeseeable and too numerous to allow enumeration in advance, require the ability to perform nonmonotonic or default reasoning in rule-based contexts. But that potentially relevant features are not exhaustively known in advance does not just affect rules and rules alone. They are equally unavailable for any mode of organizing conceptual knowledge. Thus, realistic models of categorization must allow representation and evaluation of "novel" features.

That this is not just a pedantic point that can be ignored in practice is documented by work in machine learning and artificial intelligence. The problem is well known in the context of rule-based systems (Reiter 1980), but instance-based approaches to classification in machine learning have also recognized the need to confront the problem of "novel attributes" (Aha 1992). The aim of this research is to build classification systems that work with practical problems, not ambitious cognitive models. Cognitive modeling should treat the issue all the more seriously.

There is a serious problem, then, for any account of categorization that assumes fixed representations, whether this strait-jacket of uniform representation stems from practical considerations about representation and learning procedures (e.g., backpropagation networks) or stems from the very nature of theory (e.g., spatial models of similarity).

Our own approach to similarity and categorization is based on the notion of transformation between objects, a general concept that encompasses similarity as "feature-overlap" or as distance in similarity-space as a special, restrictive case (Chater & Hahn 1997; Hahn & Chater 1997). Similarity between objects is assumed to depend on the ease of transformation of the representation of one object into representations of the other. Psychology has seen transformational accounts of similarity advanced in the past (Franks & Bransford 1971; Imai 1977). Our account of "representational distortion" provides a foundation in terms of the notion of Information Distance from the branch of algorithmic complexity theory known as Kolmogorov complexity (Li & Vitanyi 1993).

Crucial for the present context is the concept that similarity assessment no longer conceives of objects as residing in a feature space, but instead in transformation space. Features are only of interest as the objects of transformations; in this sense, the account is independent of particular features. As a consequence, there is no need for the same set of features to be present throughout. Also, the same features can be the object of different transformations as these arise from the particular pair of stimuli under consideration. The search for transformations itself influences the features found; consequently, the same basic features can give rise to different stimulus descriptions as a function of the particular comparison.

A framework for structural constraints on feature creation

Scott A. Huettel and Gregory R. Lockhead

Department of Psychology: Experimental, Duke University, Durham, NC 27708. **huettel@psych.duke.edu**

Abstract: We address two major limitations of Schyns et al. First, we clarify their concept of "features" by postulating several levels for processing. The composition of the feature set at each level determines the set at the next higher level, following simple structural guidelines. Second, we show that our proposed framework reconciles feature-creation and fixed-feature approaches.

Cognitive psychology frequently suffers from a longing for reduction. Researchers approach problems with a desire to atomize complex behaviors, seeking basic units underlying a resultant process. Schyns, Goldstone & Thibaut deny that such reductionism is possible, at least in categorization processes. They claim that this goal – to establish a fixed set of basic units governing all of categorization and object recognition – cannot be realized; people create new features as needed, and if possible, to meet the functional demands of the task.

For Schyns et al., a feature is "any elementary property of a distal stimulus that is an element of cognition, an atom of psychological processing" (sect. 1). This definition implies that there are well-defined basic units of psychological processing, which is the very position they wish to refute! We propose another interpretation: that there are several levels of psychological processing, each with its own basic units. Largely consistent with Schyns et al., we discuss a classification task in which these multiple levels may be discerned. Consider Figure 1, where the participant's goal is to articulate a simple rule that distinguishes the members of the left column from those in the right. Simple structures, such as geometric shapes, colors, and positions, are readily apparent. These units of basic structure are analogous to the *primitive features* discussed by the authors and represent a first level of psychological processing. A second level of processing corresponds to properties of a given exemplar of a category; the upper-left item has the property of "occlusion." We state that each such property is a *candidate feature,* because it presents a possible rule for categorical division. Third, and finally for such tasks, properties of the entire set of stimuli, *categorization features,* define the rule by which the classes are distinguished. For the stimulus set in Figure 1, white occluding vs. black occluding may be a basis for categorization.

From this perspective, feature creation has a well-defined meaning: something becomes available for psychological processing that was not present before. Schyns et al. seem to argue that a feature creation process will be invoked whenever the previously selected set of features is inadequate for the task. Here, features available at one level provide information to the next. When primitive features lead to a candidate feature which, in turn, is accepted as a categorization feature, feature creation is unnecessary. In contrast, when the initially generated lower-level features do not lead to useful features at the next higher level, the system returns to an earlier level of processing and changes the set of features to be submitted to the subsequent level. By this view, feature creation does not necessarily occur and should not be restricted to categorization-level features; lower-level feature sets may be changed as a result of higher-level processes.

Schyns et al. present evidence that feature creation does occur, and their Table 1 lists characteristics that encourage feature creation, but why this happens in some settings but not others is less evident in their analysis. The Martian cell provides a suitable example: the stimuli used have all the attributes of their suggested alternative materials. If experimental subjects were challenged to sort stimuli of this type in a free classification task, they would do so in a manner indicative of feature creation. Imagine if this stimulus set were changed in one way: if half the stimuli were colored blue and half were colored white. Now, subjects might sort these stimuli without recourse to the properties generated by a

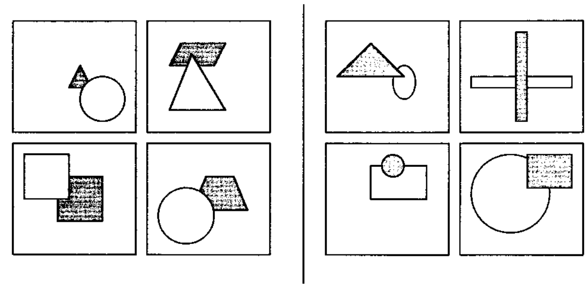

Figure 1 (Huettel & Lockhead). A sample classification problem (details in text). Adapted from Bongard (1970).

feature creation process, due to the ease of sorting along the simple color dimension. Thus, two stimulus sets, with nearly all of the same characteristics, may provide conflicting evidence for fixed-feature and flexible-feature approaches.

A resolution of this apparent conflict may be found in discussions of the role of context in the identification of stimulus dimensions (Garner 1962; 1974; Pomerantz & Lockhead 1991, especially pp. 1–14). If there is no detectable, systematic variation along a stimulus dimension, then that dimension may not be processed further. For example, the simple shapes in Figure 1 vary greatly in size, but that variation is not systematic, and so size is not immediately perceived by the subject as a candidate feature for classification. Similarly, all the items are presented on a colorless ground that could theoretically be perceived. However, without any variation of the ground, the dimension of "ground color" is effectively nonexistent. Dimensions with detectable, systematic variation across members of the set generate features to be processed, whereas dimensions that do not vary are not processed further.

This dropping out or nonselection of features based on characteristics of the total stimulus set pares down the number of features at one level that may be analyzed at the next level. For the Martian cell example, coloring some of the items blue enables the subject to detect easily that systematic variation in the stimulus set. Without such coloring (or some other highly distinctive change), the problem of detecting variations and testing them as candidates toward categorization features reemerges. Furthermore, were there color variations that did not contribute functionally to classification, so these variations will hamper detection of less apparent, but functionally discriminating, features.

This analysis of multiple levels reconciles the fixed-feature process often indicated by traditional materials and tasks with the flexible-feature perspective suggested by less-common approaches. In many studies, the information that is readily available at one level is adequate for processing at the next. This occurs when variation in the stimulus set is systematic and detectable, allowing the organism to step smartly through the levels without recourse to feature creation. However, when features at one level do not provide useful information at the next level, as when a candidate feature does not distinguish between categories, the organism must revise the set of features taken from the earlier level, creating new elements for processing.

In many real-world situations, not all information concerning the functional categories may be available. Subjects must then use existing internal representations (e.g., inferred subsets), expectancies (e.g., sources of attention), and surely much more, to help in classifying. Thus, the classification process is dynamic: what occurs depends on the structure of the stimulus set and on the observer's expectations.

Finding the Pope in the pizza: Abstract invariants and cognitive constraints on perceptual learning

John E. Hummel and Philip J. Kellman

Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095-1563. **jhummel@psych.ucla.edu; kellman@psych.ucla.edu**

Abstract: Schyns, Goldstone & Thibaut argue that categorization experience results in the learning of new perceptual features that are not derivable from the learner's existing feature set. We explore the meaning and implications of this "nonderivability" claim and relate it to the question of whether perceptual invariants are learnable, and if so, what might be entailed in learning them.

Schyns, Goldstone & Thibaut argue that visual features governing object classifications can be created by categorization experience. This is an important idea (if not a completely novel one; see, e.g., Biederman & Schiffrar 1987), but the key unanswered question is where the new features come from.

Schyns et al. argue extensively that the features created during category learning are "not present in, or derivable from, the [existing] feature set (sect. 1.1, para. 4)." It is easy to understand why they would want to make this claim (henceforth, the *nonderivability* claim): if the features acquired during category learning are just concatenations (intersections and unions, or worse – simple weighted sums) of features that existed before category learning, then the phenomenon of "category-based feature learning" might be construed as a simple matter of "selection" or "weighting." Although Schyns et al. treat nonderivability as a stepping stone to the broader claim that category learning constrains feature perception, the issue of nonderivability is arguably the more important issue. Among other things, it relates to the notion of abstract *invariants,* which are important in shape perception and object recognition (Biederman 1987). Insight into whether and how invariants can be learned from experience would make a substantial contribution to our understanding of object perception, recognition, and categorization.

Clearer understanding of nonderivability is necessary in tackling this question. There are at least three senses in which some new feature might be nonderivable from the population of previously existing features in the system (although it is unclear which Schyns et al. intend). The most literal interpretation is that the new feature is not derivable (computable) *at all* from the existing features. This version of the claim is absurd: any feature that is detected by the visual system must be computed from some finite set of operations performed on the representations given by early visual processes.

The second and third interpretations of nonderivability are more interesting because they characterize, respectively, two different ways of detecting features in any computational system. The second interpretation is that the new feature is not a simple weighted sum (i.e., linear combination) of existing features. This interpretation is suggested by the discussion of XOR, a famous example of a function that cannot be computed by a linear system. On this version of the claim, the perceptual/categorization system can be viewed as analogous to a large, multilayered neural network whose units have nonlinear activation functions (such as a standard backpropagation net). New features in one layer of the system would be composed in a nonlinear way (e.g., as a weighted sum subjected to a threshold) from existing features. This approach to feature detection and learning is standard fare in artificial neural networks. On this interpretation of the nonderivability claim, Schyns et al.'s theory is (essentially) that category learning (e.g., in "higher" layers of the network) serves to guide feature learning (in lower layers). This would be interesting, but not earth-shattering.

The third – and most interesting – interpretation of nonderivability is that the newly discovered feature is an abstract invariant, which, although *computable from,* is not truly *definable in* the vocabulary of existing features. For example, no logical concatenation – conjunctive, disjunctive, or otherwise – of local retinal activations defines the invariant square. It is acordingly a mystery how the visual system discovers such invariants in the outputs of local features (such as edges). Where invariants are concerned, it is not the case that "novel visual features are certainly reducible to their retinal encodings" (sect. 2.7). The argument is similar to those arising in discussions of scientific reductionism (Putnam 1975). Put simply, squareness is both more and less than any finite set of retinal activation patterns. It is more because some new activation pattern might also be a square, and it is less because many of the attributes of retinal activation patterns have nothing to do with their squareness. "Square" is an abstract invariant. If this is what Schyns et al. mean by nonderivable, then their claim is that category learning directs the discovery of invariants, as Gibson (1969) suggested some time ago. To our knowledge, no one has demonstrated how such invariants are discovered. The question of how (and whether) nonderivables such as invariants can be learned is a computational/algorithmic one that demands a far more specific theory than the one presented in the target article.

Toward that end, it is important to appreciate that discoverable new features do not include all logically possible ones, as Schyns et al. seem to suggest. Rather, human cognition is organized (constrained) for the discovery and synthesis of overlapping patterns in space and time. For example, we are better at detecting and learning about spatial (and temporal) relationships among parts that are close together rather than widely separated, and we are much more sensitive to some kinds of shape attributes than others (compare locating first-derivative discontinuities in contours [corners] vs. third-derivative discontinuities). Some well-defined attributes are unlearnable or even undetectable (Julesz 1981). Many of the answers to the mystery of where new features come from will probably emerge from identifying constraints on the vocabulary of spatial and temporal properties and relations that make up the human endowment for perception and perceptual learning.

Can features be created on the fly?

Koen Lamberts

University of Birmingham, Birmingham BI5 2TT, England, United Kingdom. **k.lamberts@bham.ac.uk**

Abstract: It is argued that feature creation may not only depend on categorical distinctions that are made during category learning, but also on the choice set during subsequent categorization.

Schyns et al. argue convincingly that higher-level cognitive processes influence the lower-level features that are created and used. I largely agree with their analysis and with its conclusions. Although the influence of learning on low-level processes has been studied for a long time, it is good to see a systematic and rigorous exploration of the effects of category learning on feature creation. In this commentary, I will discuss one issue that has not been addressed in the target article.

My argument concerns the role of choice sets in category learning and categorization. Which features are functionally optimal will ultimately depend on two elements: (1) the categorical distinctions that need to be made during category learning, and (2) the set of category alternatives that are considered during subsequent categorization. Schyns et al. address only the first of these two elements, but I will argue that the second may be equally important if we want to understand how higher-level processes affect low-level feature creation.

Category learning in daily life differs in many respects from category learning in the typical laboratory experiment. In many category-learning experiments, there are only a few mutually exclusive alternative categories available. However, in daily life, the set of alternatives is usually much larger and often implicit.

Moreover, objects can belong to many categories at the same time. For instance, when a child learns what a "fish" is, it learns to assign objects to one particular category, at the exclusion of a potentially very large set of alternatives: a fish is not furniture, nor is it a cat or a tree. However, a fish is also a living creature, it can become food, and it can swim. Category learning usually involves acquiring the ability to assign objects to more than one category among a very large number of alternatives. Schyns et al. argue convincingly that this process will be helped tremendously by flexible feature creation.

However, this may be only part of the story of flexible features. In real-life categorization, there can be discrepancies between choice sets during category learning and choice sets in subsequent categorization. Often, the choice set in categorization is limited by the task at hand or the processing context. Which features are optimal in a given context depends on the choice set that is available. Different features will be optimal in deciding whether a given object is an apple or a pear, for example, than in deciding if it is an apple or a tennis ball. It seems reasonable to assume that decisions are often made between category alternatives that were never before considered together. One might expect a truly adaptive system to have the ability to tailor feature uses to such specific contextual demands, without extensive additional learning. The question then becomes whether features can be created on the fly to address categorization needs in different contexts. This is essentially an empirical question that should be addressed in the future. It has been shown that categorization context can affect the distribution of attention across existing features (Lamberts & Chong 1997), but it is unclear whether truly new features can be created to serve short-term task demands. If categorization context can determine feature creation in such a flexible manner (and I believe this is a distinct possibility), fixed-feature theories of categorization and identification would be in even greater difficulty than the work of Schyns et al. suggests. Such additional flexibility would also imply that features created as a result of extensive category learning may be highly volatile, and not applicable in choice contexts that differ from those in training.

New features for old: Creation or derivation?

Cyril R. Latimer

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

Abstract: Schyns, Goldstone & Thibaut oppose the notion of fixed feature analysis, suggesting the possibility of flexible feature creation in object recognition and categorisation. Such proposals cannot be assessed until clear definitions of the objects in question and their decompositions are formulated. Flexibility may come from the decompositions of objects rather than from feature creation.

Schyns, Goldstone & Thibaut raise the important and venerable issue of the way the perceptual world is analysed by our senses and the relativity of the terms "whole" and "part." What are the perceptual primitives and are they immutable, or are they flexible and created to suit particular contexts? Central to their arguments is their opposition of fixed-space versus flexible-space category learning and fixed features such as "geons" or pixels versus flexible features. The difficulty with the latter dichotomy is that fixed high-level geons may not be flexible enough to capture the representation necessary for a new task, whereas regularities at the level of functional features may not be captured by or derivable from a high-resolution, pixel decomposition of category members. One difficulty the authors face if their theory of flexible feature creation is to be testable is that before any feature, part, or dimension can be pronounced old, newly created, derivable, or underivable, they themselves must specify some initial set of inflexible primitives.

Wholes, parts and derivable/underivable properties. The difficulty in discussing wholes and their parts was alluded to in a seminal paper on the analysis of gestalt concepts by Rescher and Oppenheim (1955). Those authors pointed out that because wholes and their parts may have an indefinite number of descriptions and decompositions, there can be no talk about any whole and its parts until the whole in question is defined precisely and in enough detail. Moreover, the decomposition of the whole into parts must also be described precisely, along with details about the putative parts and their attributes. If some nonspatial or dimensional decomposition of a whole is adopted, then the dimensions and their ranges of values and so on must be precisely defined. Only then can there be coherent discussion of whether specified properties of a whole are old, new, derivable, or underivable from the specified parts of the whole and their attributes.

For example, Schyns et al. cite the example of "symmetry" as an instance of the "functionally important object regularities . . . [that] are often not captured by simple pixel-based features" (sect. 2.4.1.). Whether the property of symmetry is captured or derivable from a pixel-based decomposition depends entirely on how the decomposition and its attributes are defined. For some decompositions (e.g., Gibson's 1969 graphemic analysis, which does not specify the locations of parts within characters), symmetry will not be derivable directly from the parts. For other decompositions (Latimer et al. 1994; Sejnowski et al. 1986), symmetry will be derivable.

Theory-dependent derivability. The situation is further complicated because derivability depends not only on the decomposition and attributes chosen but also on the choice of background theory. Some theories, by virtue of their assumptions, will allow higher-order holistic properties to be derived, whereas other theories will not. Consider, for example, how features like "curvature" are encoded by the visual system. In a theory in which curvature is underivable, curvature is somehow extracted directly by curvature detectors in the visual system and is not derived from the products of local analysis (Riggs 1973; 1974). In a theory in which it is derivable, large curved lines may be registered by cells whose optimum stimulus is a straight line. Thus, an arc may be derived from the outputs of short line detectors located so that they are tangents to the larger arc. As it happens, the weight of evidence supports this theory (Blakemore & Over 1974; Crassini & Over 1975).

Conceptual and empirical implications. Schyns et al. admit that it will be very difficult to determine whether experimental participants are creating, deriving, or simply reweighing object features. Leaving aside this empirical issue for the moment, the conceptual implications of the foregoing remarks for Schyns et al.'s notion of feature creation are clear. In any investigation of whole/part perception, it is first necessary to specify in sufficient detail (1) the properties of the wholes to be recognised or categorised, (2) the chosen decomposition of the wholes and their attributes, plus (3) a detailed description of any relevant background theory. Only then will it be possible to determine by logical analysis and simulation whether or not properties are derivable. Objects with features shown to be new and underivable in simulations could be tested on experimental participants, but the difficulty, I suspect, will lie in finding such underivable features. In the myriad possible sets of objects, decompositions, attributes, and associated background theories, where are such features to be found? For every decomposition that does not allow a particular property to be derived, it may be possible to find another that does.

New features or new decompositions? Instead of reweighing, deriving, or creating new features, perhaps the brain adopts new and better decompositions that allow diagnostic features to be derived. For example, Pasnak (1971) found that although a three-by-three decomposition proved to be predictive of the time course and errors in same/different judgments about complex geometric forms, the same decomposition did not explain the results of experiments with simpler forms. Thus, the flexibility proposed by Schyns et al. may lie in the brain's ability to adopt efficient and context-sensitive decompositions rather than having to create or derive features from a fixed decomposition? In any event, the conceptual and experimental analysis of part/whole perception must confront the conceptual and definitional issues described here. The importance and strength of Schyns et al.'s contribution lie in highlighting the currency of these conceptual matters in both object recognition and categorisation.

Feature learning, multiresolution analysis, and symbol grounding

Karl F. MacDorman
Department of Mechanical Engineering, Osaka University, Toyonaka, Osaka 560, Japan. **karl.macdorman@cl.cam.ac.uk** **www.cl.cam.ac.uk/~kfm11**

Abstract: Cognitive theories based on a fixed feature set suffer from frame and symbol grounding problems. Flexible features and other empirically acquired constraints (e.g., analog-to-analog mappings) provide a framework for letting extrinsic relations influence symbol manipulation. By offering a biologically plausible basis for feature learning, nonorthogonal multiresolution analysis and dimensionality reduction, informed by functional constraints, may contribute to a solution to the symbol grounding problem.

Schyns et al. present compelling arguments (sect. 2.8) and evidence for the ability to learn diagnostic features. It is accordingly useful to begin by considering why fixed features have proven so popular. Perhaps most important, fixed features provide a straightforward basis for framing cognitive theories (sect. 1.1). Fixed features and combination rules enable a symbol system to simulate key aspects of human thinking such as its systematicness, productivity, and semantic coherence. Fixed features also simplify the mechanics of deduction and abstract planning by separating the symbol system from the details of sensorimotor control.

Trouble arises because, in current systems, symbol manipulation turns on properties *intrinsic* to the system (i.e., syntactic constraints and the physical properties of the implementation media). However, even proponents of symbol systems (Fodor 1994) now admit that *extrinsic* relations influence intentional contents – in particular, the causal relation between a thought's content and what it represents. To ignore this fact leads to the symbol grounding problem (Harnad 1990).

Standard artificial intelligence solutions fail because (1) outside of simple domains, a programmer cannot anticipate all necessary elementary features and, hence, cannot set up a priori feature detectors; (2) to represent all sensorimotor information symbolically creates unreasonable computational demands (Janlert 1996); and (3) analog information needs to bear on abstract reasoning. Otherwise, symbol systems lack empirical constraints and (having only formal constraints) can define a limitless number of "kooky" concepts (Fodor 1987). This excess of freedom contributes to the frame problem. Its solution requires finding a representational form that obviates the need to reason about stabilities (Janlert 1996). However, our best hope for that is precisely the bottom-up perceptual and functional constraints lacked by current symbol systems (Harnad 1993). This is one reason why we may need representational forms that fall along an analog-categorical continuum (Harnad 1987).

By letting extrinsic relations (as mediated by the body) influence internal symbol manipulation, feature learning offers a more sound foundation for cognitive theories. Feedback from failed sensorimotor predictions could form the basis for learning analog-to-categorical mappings (MacDorman 1997a; 1997b). These could in turn ground symbols and rules. Sensorimotor predictions correlate the sensory and internal consequences of motor activity. When they are violated, orienting reactions ensue that guide attention toward the source of error. These predictions underlie the ability to distinguish object affordances. For this reason, they must integrate information from multiple modalities. This provides a framework for more passive or abstract feature learning. Active sensorimotor predictions constitute an individual's current world-model.

More specific issues are discussed in the following sections.

Viewpoint. There is good reason to doubt the existence of a unique task-independent view-based representation of an object (sect. 2.1). One alternative is to learn a viewpoint-independent representation of an object by developing predictions concerning how bodily movements transform the object's (viewpoint-dependent) sensory projections (MacDorman 1997a; 1997b). Using these, it is possible to shift from a viewer-centered to an object-centered frame of reference, because predictions concerning how self-induced movements cause sensory transformations can be used to compensate for those movements. [See also, Edelman "Representation is Representation of Similarities" BBS 21(3) 1998.]

Nonorthogonality. As Schyns et al. point out, the categorization process can profitably inform the dimensionality reduction of its input (sect. 3.4.2). Although this approach makes a featural interpretation of principal components easier, one drawback is the high cost of each recalculation of the eigenvectors used by the Karhunen-Loève transform. Another drawback is that biological sensorimotor systems use *non*orthogonal coordinates. Orthogonality would exact a high cost both genetically and in terms of neural development. It leaves visual processing susceptible to changes in receptive field profiles. If the brain used orthogonal coordinates, neural death could easily render some patterns imperceptible and others indistinguishable.

Multiresolution analysis. Results in character and face recognition research from using wavelets to pursue input signals at various scales support Schyns et al.'s conclusion that larger features need not be composed from smaller ones (sect. 3.1). Daugman (1980) proposed that a parametrized family of two-dimensional Gabor filters (nonorthogonal wavelets) offers a suitable model of the anisotropic receptive field profiles of single neurons in several areas of the primary visual cortex. The two-dimensional Gabor filters are able to account for the selective tuning of simple cells for characteristic scale, localization, orientation, and quadrature phase relationships. Daugman (1985) has shown using chi-squared tests that this family of elementary functions fits the profiles of 97% of simple cells in the cat visual cortex. Within-category invariance (with between-category invariance filtered out) learned from the output of Gabor filters may underpin flexible features (MacDorman 1997b).

Schyns et al.'s use of alternative materials is laudable. Virtual reality may soon permit us to pursue a *truly* bottom-up multimodal investigation of feature learning. Subjects possessed of "bodies" with utterly different kinematics, sensors, and actuators will move in alien perceptual worlds, and we will get to study them.

Formal models and feature creation

Thomas J. Palmeri
Department of Psychology, Vanderbilt University, Nashville, TN 37240.
palmerit@ctrvax.vanderbilt.edu
www.vanderbilt.edu/AnS/psychology/cogsci/palmeri/home.html

Abstract: Formal models provide detailed accounts of fundamental aspects of categorization, yet are potentially incomplete in not providing accounts of how new features might be created. Although the notion of feature creation is enticing, how this complex process might operate is not specified. Moreover, arguments favoring feature creation accounts that are founded on the implausibility of feature weighting accounts are not convincing.

Schyns, Goldstone & Thibaut provide some compelling arguments in favor of flexible feature creation as opposed to the fixed feature spaces assumed by many theories of categorization and object recognition. I agree that although assuming fixed feature spaces should not be problematic for categorizing and recognizing fairly simple experimental stimuli, such as forms of canonical shapes and colors, flexible features may be necessary for categorizing and recognizing more complex stimuli, such as faces or radiological images. Although I find many of the ideas presented in the target article enticing, this commentary will instead focus on some misconceptions, mischaracterizations, and omissions as they pertain to categorization.

Formal models and simple stimuli. Researchers interested in developing formal computational models of categorization have often used stimuli that "wear their features on their sleeves." As a result, they have largely overlooked the role of emergent features in category learning. Why? Is it because formal modelers believe that all stimuli are composed of such manifestly present features? Certainly some might hold this theoretical viewpoint. However, many modelers use stimuli such a simple shapes, colors, or tones *precisely because emergent dimensions do not form,* not because they believe that emergent dimensions are inconsequential. In experiments using such stimuli, high levels of control over features are necessary to evaluate detailed aspects of some proposed models.

It is certainly true that "feature creation" has been ignored by most formal models of categorization. This is largely because the mechanisms involved are not very well understood and pertinent empirical work is lacking. This target article should provide some impetus for researchers to take these issues seriously in future work. It must be emphasized, however, that a tremendous amount has been learned about detailed aspects of categorization using simple stimuli (e.g., see Ashby 1992; Estes 1994; Nosofsky 1992). Although it is difficult to believe that Schyns et al. would deny this, they unfortunately minimize the insights that have been gained through formal approaches when they claim that the "lion's share" of the work in much of real-life category learning is in feature creation. Because they present relatively limited empirical evidence and little concrete theoretical development, I simply am not convinced that most of the "work" in category learning, by whatever metric this might be assessed, is done by feature creation.

Although specifying how appropriate feature spaces might be created is without question an important unresolved issue, specifying representations is insufficient for a complete account of categorization. Feature creation is but one (albeit largely ignored) aspect of categorization. Even if new features are created (by some presently unspecified mechanism), issues of how these features are used to make a categorization decision must still be addressed: How are the features of a presented object compared with information stored in memory? What information about a category is stored in memory? What kind of categorization decision process is used? Formal models have provided some answers to these important questions.[1] I would be surprised if the insights gained using simple stimuli were found to be of little value when "alternative materials" are used.[2]

Implausibility arguments and constraints on feature creation. One of the greatest challenges to a feature creation account is to distinguish it from a feature weighting account. Although I find many of the arguments Schyns et al. raise against a feature weighting account sensible, too often their appeals are grounded in "implausibility." Feature creation is necessary if "the required number of prespecified features would otherwise be implausibly large" (sect. 1.2), and positing a feature weighting account would require "positing an implausibly large number of dimensions or features" (sect. 1.2.2, para. 3). "Implausibility" is a fairly weak foundation for theoretical arguments, especially when the basis for such an assessment is left unspecified.

Exemplar-based models of categorization were once deemed implausible. Given a pervasive computer metaphor for the mind, memory capacity was limited and processing was serial, ruling out the possibility of storing and retrieving specific category instances. Abstractionist theories of categorization hence dominated the field. It is now well recognized, however, that processing in the brain is highly parallel and storage capacity is vast. Although plausibility arguments are still raised, exemplar-based theories of categorization, object recognition, and other cognitive processes are now some of the most successful in the field.

Schyns et al.'s appeals to "implausibility" are not very convincing. In the case of feature weighting, we have a fairly simple selective attention mechanism operating over a potentially large feature space.[3] It is not that difficult to imagine how such a mechanism might operate (e.g., Kruschke 1992). In the case of feature creation, we have a complex mechanism operating to create a simple feature space. How this mechanism might work is unknown. However, it must be powerful enough to create efficient, specially tailored representations that distinguish between object categories. Although the kinds of features that can be created must be constrained by perceptual processes, the nature of these constraints is unspecified.

Which account is more plausible depends on one's metric of complexity. Whereas Schyns et al. see the sheer size of the feature space as a limitation of a feature weighting account, I see the sheer complexity of feature creation as a potential limitation of a feature creation account. In my view, until a testable model of feature creation is specified in some detail, the plausibility of such an intelligent feature creation agent must remain suspect instead.

NOTES

1. As for one successful formal model, Schyns et al. mischaracterize RULEX (Nosofsky et al. 1994) when they claim it assumes that people learn categories "by discovering complex rules that integrate several distinct stimulus features" (sect. 2.8, para. 5). Rather, RULEX assumes that people form simple rules along one, or possibly two, psychological dimensions and then remember exceptions to those rules. In fact, RULEX could be a useful decision mechanism to supplement the purported feature creation process – unless a single necessary and sufficient created feature signals category membership, exceptions are a likely possibility.

2. Schyns et al. omitted any discussion of the vast categorization literature using random dot patterns (e.g., Homa 1984; Posner & Keele 1968; Shin & Nosofsky 1992) – clear examples of "alternative materials." It is true that multidimensional scaling methods are often necessary to derive the psychological dimensions of these patterns. Yet, when these derived configurations are assumed, formal models have provided excellent accounts of categorization performance (e.g., Shin & Nosofsky 1992; Palmeri & Nosofsky, submitted). I agree that origins of the psychological dimensions are left unspecified by current models. However, specifying these origins does not provide a theory of categorization in itself.

3. Schyns et al. mischaracterize feature weighting accounts in general as a form of curve fitting (sect. 2.5). In fact, Kruschke (1992) has developed a connectionist model that learns to attend to diagnostic dimensions, and Nosofsky (1984) has cast feature weighting in terms of optimality considerations.

Feature creation as a byproduct of attentional processing

Pierre Perruchet and Annie Vinter

LEAD, Faculté des Sciences, Université de Bourgogne, 21000 Dijon, France. **perruche@u-borugogne.fr; vinter@u-bourgogne.fr**

Abstract: Attributing the creation of new features to categorization requirements implies that the exemplars displayed are correctly assigned to their category. This constraint limits the scope of Schyns et al.'s proposal to supervised learning. We present data suggesting that this constraint is unwarranted and we argue that feature creation is better thought of as a byproduct of the attentional, on-line processing of incoming information.

In traditional category learning studies, new categories emerge from new combinations of a fixed repertoire of elementary features. Schyns et al. show cogently that low-level features can themselves change with experience, thus altering the immediate appearance of objects. We fully agree with this perspective, which amplifies the impact of learning to the deepest roots of perception and categorization, thus converging with other sources of evidence that the role of learning in cognition has been underestimated (e.g., Bates & Elman 1996, with regard to language).

We believe, however, that one aspect of the Schyns et al. proposal may ultimately limit its implications. They claim repeatedly that feature creation depends on categorization requirements, and that people create features to subserve the representation and categorization of objects. We have no special quarrel with this proposal insofar as it intends to describe the ultimate function of features in adaptive behavior. However, the authors mean something much stronger, namely, that the categorization requirement is the actual driving force in extracting distinctive features. The difference is crucial, as it appears when comparing the hardly disputable claim that mating occurs in the service of species survival, and the contention that individual sexual behavior is initiated and shaped by this ultimate function. Conceiving categorization requirement as the actual causal agent for feature creation undermines the Schyns et al. model.

The restriction to supervised learning. The problem is that the tightly functionalist stance by Schyns et al. limits the relevance of their model to situations in which participants are informed about the nature of the categories. Indeed, in the componential view of cognition to which Schyns et al. seem to subscribe, categories are themselves defined by their distinctive features. It would obviously be circular to simultaneously ground category formation in feature knowledge and feature creation in category knowledge. To avoid circularity, the claim that features are learned insofar as they are needed to achieve categorization requires that displayed exemplars be correctly assigned to their categories by an external informer. Although Schyns et al. allude briefly to the beneficial effects of preexposure without external feedback (sect. 1.2.1; see also the preliminary experiments by Schyns & Rodet 1997), they are aware of this constraint. To quote them, "the individual knows what the categories are from external feedback" (sect. 2.2).

Because feature creation, according to Schyns et al., implies concurrent category knowledge, extending the scope of learning to features paradoxically prevents genuine category discovery. We are then faced with the following alternative: either new categories are formed by combining known features, the conventional view, or new features can be created from the knowledge of fixed categories, as claimed by Schyns et al. Disappointingly, both sides of this alternative rule out the possibility that people faced with a new environment can learn both features and categories by themselves. We subscribe to the view that new features can be created, but we intend to show that the process is functionally independent of categorization requirements. Our proposal is that feature creation makes it possible to form new categories instead of requiring information about categories.

Evidence for feature creation in unsupervised learning. The observation that people may be able to parse complex material according to its relevant parts in unsupervised learning without any surface cues or external information about the deep structure of the material has occasionally been reported in various areas of research (e.g., see Saffran et al. 1997). However, most of the evidence comes from the so-called implicit learning studies. In these studies, subjects are typically faced with complexly structured material, such as a set of letter strings, the order of which is defined by a finite-state grammar. There is evidence that the subjective encoding units of such a complex display, which are initially randomly determined or driven by possibly irrelevant surface features, become increasingly congruent with the deep structure of the material (Perruchet & Gallego 1997; Servan-Schreiber & Anderson 1990). As in the situations described by Schyns et al., learning is responsible for the formation of the building blocks of cognition, instead of dealing with only the storage, processing, and retrieval of preshaped representations. (It may be pointed out that one deals with subjective *units* in the implicit learning context, whereas Schyns et al. deal with *features.* This difference may be terminological, insofar as most of the experimental support cited in the Schyns et al. target article discussed the segmentation of objects into parts in the same way that in implicit learning experiments the training material is segmented into perceptual units.)

The crucial difference between the data provided by Schyns et al. and the results just described lies in the fact that in the latter case, the building blocks of cognition are shaped without any external information about the categorical structure of the material.

To account for this finding, we have proposed a model that relies on simple and ubiquitous attentional and memory processes (Perruchet & Vinter 1997; Perruchet et al. 1997). The initial perceptual units composing the momentary focus of attention are determined at random, or result from various bottom-up influences such as those described by Schyns et al. (sect. 2.5). Some of these perceptual units presumably match the structurally relevant parts of the material, whereas others result from irrelevant fragmentation. The key point is that a given part tends to be repeated only when it is structurally relevant, as a mandatory consequence of the rule-governed structure of the material. This entails that irrelevant units, because they reoccur infrequently, will be forgotten, whereas the initial selection of meaningful units will be progressively reinforced by repetition. With repeated exposure, subjective units become strong enough to shape perceptual processes and alter the immediate appearance of objects. Thus, in this model, the formation of cognitive units matching the meaningful part of the material is a byproduct of the attentional, on-line processing of incoming information.

After their initial exposure to the letter strings in artificial grammar learning, subjects are able to discriminate new grammatical and ungrammatical strings. This shows that feature creation, at least when viewed as a structurally relevant parsing of the environment, can be one basis for the formation of new categories, instead of being grounded in externally induced category knowledge. Our proposal does not entail that categorization requirements are never causal factors in feature creation, as claimed by Schyns et al. However, this form of processing may be limited to the cases in which people are explicitly asked to solve problems or to engage in analytic forms of thought.

ACKNOWLEDGMENTS
Preparation of this article was supported by the University of Bourgogne and the Centre National e la Recherche Scientifique (ESA CNRS 5022).

Context-dependent feature discovery is evidence that the coordination of function is a basic cognitive capacity

W. A. Phillips
Center for Cognitive and Computational Neuroscience, University of Stirling, FK9 4LA, Scotland, United Kingdom. **wap1@forth.stir.ac.uk**
www_psych.stir.ac.uk/~wap/

Abstract: Schyns et al. make a strong case for context-dependent feature discovery. The features computed from specialized and diverse data-sets help to coordinate their activity by adapting so as to emphasize what is related across sets. Their perspective can be strengthened and extended by formal arguments for the contextual guidance of learning and processing and by neurobiological and psychological evidence of structures and processes that implement this guidance.

The specialization of function has long been central to our understanding of cognition. It is now becoming clear how this is complemented by the coordination of function. Argument and evidence that contextual guidance is a fundamental capability of cortex and cognition are reviewed by Phillips and Singer (1997), where the goal of discovering features that are predictably related to the context in which they occur is formally specified, implemented as a learning algorithm, and shown to be supported by known anatomical structures and physiological processes. The slim psychological case for context-dependent feature discovery presented in that review (sect. 5.6 & 5.7) is now greatly strengthened by Schyns et al. and particularly by their psychological evidence of learned features (target article, sect. 1.2). In addition to providing computational and neural support for the processes they propose, our perspective can: (1) help clarify their distinction between object concepts and other levels of description; (2) extend the possible sources of guidance to include within-level interactions as well as top-down influences; (3) extend the range of variables discovered to include those that are continuous as well as those that are categorical; and (4) help emphasize the role of discovered relations between the discovered variables in organizing perception.

Object concepts are distinguished by what distinguishes objects, not by nonfixedness, and soon. Although this may sound tautological, it is not, because what distinguishes objects is specified in relation to the environmental input by the classes of objects that are met, and in relation to the internal cognitive operations of object perception by the sets of stimuli to which discriminative object recognition responses generalize. Schyns et al. struggle throughout with the problem of distinguishing "object concepts" from "perception" and from "relatively raw data." Given their emphasis on learned object features, fixedness versus nonfixedness seems to be implied as one criterion, but other criteria are also suggested, for example, strategic versus nonstrategic, and the amount of prior processing required. Most of these are unconvincing. Primary sensory maps are both adaptable and affected by attention. These difficulties are unnecessary because the curse of dimensionality ensures that object concepts cannot be based directly on the sensory maps. Mediating variables are therefore necessary, and the formal argument by Schyns et al. to this effect (sect. 3.4) can be relied on. Their confidence in this argument may be increased by noting that there is also clear empirical evidence for just such a massive reduction in the information available in going from sensory storage to short-term visual memory. This selective mapping has exactly the properties required by the analysis by Schyns et al., including a loss of positional sensitivity, and thus an increase in positional generalization (Phillips 1974).

Continuous variables may also have to be discovered. The bias toward categorical variables seems to me an unnecessary restriction. When selecting a fruit to eat, it is as useful to predict its sweetness as to categorize it as an apple or a pear. An intrinsic bias toward categorical variables may prevent our discovering those that are useful but continuous. This problem can be solved by replacing the bias toward categorization with the bias toward discovering whatever variables are related to the contextual information available (Phillips & Singer 1997).

Information to guide feature discovery need not be restricted to top-down sources. I see no need to restrict the information that guides feature discovery to that provided by "top-down" sources, which are in any case not well specified. Taste information could guide the discovery of visual cues to sweetness even if the taste descriptors were computed automatically by a few stages of fixed processing. What matters is not that the guiding information comes from some high-level strategic controller but that it arose from an input that is distinct from and is (at the level concerned) kept distinct from the input whose analysis it guides.

Relations between the features must also be discovered and used. Schyns et al. imply this but provide no explicit account of how feature discovery can be combined with the discovery of the relations between those features. They do discuss their use in organizing part–whole relations in perception, however, and their arguments to this effect are supported by evidence of the role of learned contextual interactions in grouping via synchronization (Singer 1995). This raises an important and as yet unresolved problem, however, which is to reconcile the emphasis on object concepts as feature lists with the evidence for object concepts as structural descriptions. Although Schyns et al. leave this and other mysteries of object perception unresolved, the evidence of and arguments for an open set of discovered features are very likely to be a crucial part of the story.

Emergence of object representations in young infants: Corroborating findings and a challenge for the feature creation approach

Paul C. Quinn

Department of Psychology, Washington & Jefferson College, Washington, PA 15301. **pquinn@vms.cis.pitt.edu**

Abstract: Arguments for feature creation receive support from studies of young infants forming category representations from perceptual experience. A challenge for Schyns et al. will be to determine how a feature creation system might interface with a perceptual system that appears constrained to follow organizational principles that specify how edge segments should be grouped into functional units of coherent object representations.

Schyns et al. argue for a perceptual-conceptual system that is anchored by representations of object categories. These representations emerge as features are perceptually learned from visual input during the developmental course of object recognition and categorization. The major purposes of this commentary are to (1) review evidence on the development of category representations by *young infants* that provides additional support for Schyns et al. and (2) highlight one major theoretical and empirical challenge for the feature-creation approach.

Young infants are an important subject population for Schyns et al. By studying young infants, investigators gain insight into the initial beginnings of a feature-creation system. Young infant subjects minimize the problem of uncontrolled levels of perceptual experience and conceptual knowledge that are present in attempts to study feature creation by adults. Empirical investigations of object recognition and category formation conducted with young infants also provide a better understanding of the relative roles of low-level perceptual structures versus high-level conceptual theories in the early development of object parsing. The more that young infants can extract structure from "meaningless" patterns, the greater the role of perceptual mechanisms in discovering functional units of processing from surface appearance.

One idea embedded in the Schyns et al. framework is that features that come to define categories are extracted during the task of category learning. It follows that in different categorization tasks, different features will be extracted, and hence different category representations will be formed. One parameter of interest is whether a category is learned in isolation or in the context of contrasting categories. Schyns et al. predict that feature extraction may change as a category goes from being presented in isolation to being presented in the context of other categories. In particular, during single-category acquisition, an observer would be expected to register features that were *common* to exemplars of the target category. In contrast, during multiple-category acquisition, an observer should register those features that were common to exemplars from the target category, *and* that also served to *distinguish* the target category from contrast categories.

Consistent with these predictions is the performance of 3- to 4-month-olds, who develop more robust and more tightly tuned representations for form category information when presented with dot-pattern squares and triangles together than when presented with either category singly (Quinn 1987). Also relevant are findings that 3- and 4-month-olds form a category representation of domestic cats that includes female lions when familiarized with pictorial instances of domestic cats (Eimas & Quinn 1994). However, the infants form a category representation for domestic cats that excludes female lions after a complex familiarization procedure that presents pairs of cats and pairs that include a cat and a female lion (Eimas et al. 1994). Apparently information (presumably featural) about how exemplars are alike *and* how they differ from exemplars of other categories may contribute to the formation of differentiated category representations that are more nearly at the basic level in their exclusiveness. Such evidence provides support for the idea that the specific features extracted from stimuli are dependent on the nature of the categorization task presented to human observers from the very beginnings of development. Moreover, considering the age group tested and the type of stimuli used (static pictures of exemplars), at least some of the feature creation that Schyns et al. argue for is likely to be an inherent characteristic of the manner in which our cognitive systems learn information from perceptual experience.

A remaining challenge for Schyns et al. is to determine whether a system of feature creation replaces, reorganizes, exists in parallel, or works interactively with any system of object recognition that might be in place before the onset of environmental experience (e.g., Biederman 1987). For example, Gestalt psychologists have long contended that our perceptual systems come preconstructed to follow certain principles (e.g., closure, common movement, good continuation, proximity, similarity) that specify how surface fragments of a visual scene should be spontaneously grouped into more complex structures (i.e., units of processing) that serve as the basis for object recognition (Helson 1933). Recent evidence indicates that young infants adhere to at least some of these principles when organizing visual pattern information (Quinn et al., in press; Quinn et al. 1993).

To determine the interplay between adherence to Gestalt organizational principles and flexible feature creation, experiments like those conducted by Pevtzow and Goldstone (1994) should be performed with young infants. Will features that are specified as functional by Gestalt principles be "overlooked" by young infants if alternative means of object parsing are "suggested" by presenting a category of objects in which the features uniting the objects are "nonnatural" in the Gestalt sense? Such experiments should prove valuable in understanding the initial power of a flexible feature-creation system and whether it can override perceptual parsing principles that a system may come preequipped with.

To summarize, the arguments by Schyns et al. for feature creation are consistent with data on young infants forming category representations from perceptual experience. An important task for Schyns et al. will be to determine how a feature creation system might interface with a perceptual system that appears constrained to follow principles that specify how edge segments should be organized to form functional units of coherent object representations.

Parts of visual shape as primitives for categorization

Manish Singh[a] and Barbara Landau[b]

[a]*Department of Cognitive Sciences, University of California, Irvine, CA 92697;* [b]*Department of Psychology, University of Delaware, Newark, DE 19716.* **manishsi@aris.ss.uci.edu**
aris.ss.uci.edu/cogsci/personnel/astudents/singh/
blandau@udel.edu www.udel.edu/psych/fingerle/landau/htm

Abstract: Converging psychophysical evidence suggests that the human visual system parses shapes into component parts for the purposes of object recognition. We examine the Schyns et al. claim of "creation" of features in light of recent work on part-based representations of visual shape, particularly the perceptual rules that human vision uses to parse shapes.

Schyns et al.'s target article presents theoretical arguments and empirical evidence for the idea that, to characterize human concept learning, one must allow for the "creation" of new features. We believe that this thesis is important, both in the context of adult categorization and in its early development. Recent studies have emphasized the flexibility of human categorization (see Medin & Coley, in press, for a review) and the importance of considering task requirements in determining how people compute similarity relationships between objects. There is no doubt that adults are supremely flexible, using a variety of principles to formulate new categories. A consideration of how categories are learned by young children, however, forces at least an equal focus on those categorizations that are learned early, easily, and with little formal tutoring – those that are natural. Although we believe that the work of Schyns et al. is crucially important in discovering the range of flexibility in both early and mature categorization, we also believe that certain kinds of part-based categories may provide initial constraints on which learning can build.

We examine here the claim of "creation" of features in the context of part-based representations of visual shape. The issue of shape representation is extremely relevant to categorization (especially in the context of naming), and it is one to which Schyns et al. return a number of times in their article. Indeed, there is now extensive work on how shape representations are recruited during word learning in children (Landau 1994; Landau et al. 1988). Converging experimental evidence with adults shows that human vision does parse shapes into parts (Biederman 1987; Braunstein et al. 1989), that it does so quickly (Hoffman & Singh 1997), and perhaps preattentively (Driver & Baylis 1995). Part-based representations may be the initial natural units on which later learning rests.

How does human vision parse shapes into parts? Many part-based schemes of shape representation use a predefined set of shape primitives – for example, Biederman's (1987) *geons,* or Marr's (1982) *generalized cylinders.* According to these approaches, human vision parses shapes into parts by looking for these primitives in images. As Schyns et al. point out, however, this greatly restricts the scope of shapes that can be recognized, because part shapes that are not in the predefined set of primitives cannot be represented adequately. The *minima rule* (Hoffman & Richards 1984), on the other hand, makes no prior assumptions about the shapes that human vision will encounter. Instead, it uses a general computational scheme, based on the geometry of shapes alone, to find the boundaries between parts on any given shape. On a silhouette, the minima rule gives negative minima of curvature on the contour of the silhouette as boundary points. And on a three-dimensional shape, the minima rule gives loci of negative minima of the principal curvatures on the surface of the shape as boundary curves.

How does the *functionality principle* of Schyns et al. relate to the minima rule? The authors cite compelling experimental results (e.g., from Schyns & Murphy 1994; Schyns & Rodet 1997) showing that subjects with different categorization histories indeed parse shapes differently: depending on the categorization tasks that subjects have had to perform, they divide the same shapes differently into (task-relevant) features. One should note, however, that neither of the parsings produced actually contradicts the minima rule. (Such a contradiction is hinted at in para. 3 of sect. 1.2.2 of the target article, and in Schyns & Murphy 1994.) The parts delineated by the subjects on the Martian rocks, for example, are in fact bounded by loci of negative minima of the principal curvatures (and not, for example, by parabolic curves or loci of positive maxima). Subjects seem to differ in their parsings only to the extent that some of them delineate as two parts what others consider a single part – because such a distinction was important to the cateogorization task they performed.

The preceding argument does not undermine the role of the functionality principle, however. For example, it would be incorrect to say that the functionality principle defines perceptual units simply by taking task-relevant compositions of parts defined by the minima rule. This would be incorrect, because the minima rule does not define parts. It defines only *part boundaries.* It constrains, but does not define, *part cuts.* So, clearly, there have to be constraints, above and beyond the minima rule, that determine subjects' parsings of shapes.

Of course, there are many *perceptual constraints* beyond the minima rule that affect one's parsing preference. For example, the parsing produced by drawing horizontal cuts looks extremely unnatural in Figure 1a, but not in Figure 1b. Thus, relative *cut-length* and *good-continuation* are examples of perceptual constraints on part cuts. Singh et al. (1996) have recently extended the

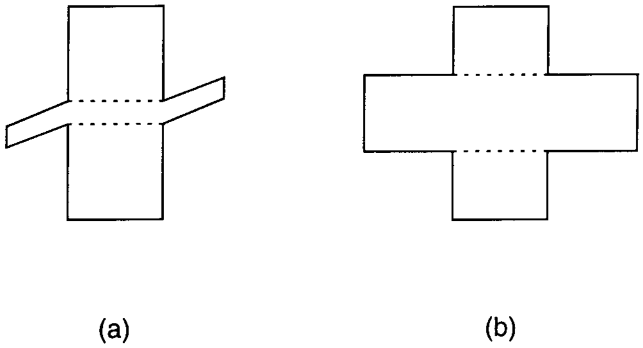

Figure 1 (Singh & Landau). Perceptual constraints beyond the minima rule affect one's parsing preference. The parsing produced by horizontal cuts looks unnatural in (a), but not in (b).

minima rule to the *part-cut rule* by incorporating more global properties of shapes into it. Furthermore, Hoffman and Singh (1997) have proposed a theory of part salience that lists various geometrical factors that make some parts more salient than others. In experimental tests of this theory, they found that their salience factors can affect the choice of figure and ground in very rapid (50 msec) presentations.

All of these factors, however, still leave room for the functionality principle to operate – although it is certainly more constrained than Schyns et al.'s example of the silhouette with 20 bumps might suggest (sect. 2.5, para. 5). The functionality principle still has room to operate because: (1) all of these factors put together still do not define a unique parsing *in all cases;* (2) these factors do not define a *hierarchy* of parts – in which large parts are seen as having further part-structures; and (3) some factors in the part-cut rule are more "flexible" than others. For example, in experiments with cross-shaped and elbow-shaped stimuli, Seyranian et al. (1997) found that although subjects consistently preferred shorter part cuts to longer ones, they were divided on how they used the "area" factor: some subjects preferred cuts that produced smaller parts, whereas others preferred cuts that produced larger parts. Studying the interactions between top-down constraints imposed by the functionality principle and such factors as salience and the part-cut rule is an extremely interesting empirical direction to pursue. Understanding the developmental interactions among these may be critical to explaining how the perceptual system comes to support categorization, and vice versa.

In sum, the case of part-based representations of visual shape suggests that one need not choose between the two extreme alternatives: one in which perceptual processes give rise to fixed features, which may then be combined in various ways to give high-level conceptual features; and the other in which high-level conceptual features are created completely anew in response to categorization demands. Instead, one may simply think of perceptual processes as imposing various constraints on the organization of the incoming data. Some of these constraints are extremely strong and rigid – perhaps innate – whereas others are more flexible and may be structured by differences in learning history. In many cases (certainly the kinds of cases that Schyns et al. are studying), all these constraints put together do not determine a unique perceptual organization. This leaves plenty of room for categorization and other task requirements to mold the formation of "features." As a result, different task requirements can lead to the formation of different features, and this certainly gives them a character of "newness."

Parts, features, and expertise

James Tanaka

Department of Psychology, Oberlin College, Oberlin, OH 44704.
tanaka@occs.cs.oberlin.edu www.oberlin.edu/~psych/viscoglab.html

Abstract: Research in expert categorization is consistent with the Schyns et al. claim that functional features are determined by constraints imposed by the learning history of the categorizer and the demands of the categorization task. However, the expertise work also suggests that a distinction should be drawn between the categorizer's perceptions of the constituent parts of the object and its functional features. Although experts and novices may parse a domain-specific object into the same parts, their featural interpretations of those parts may differ significantly.

In their target article, Schyns, Goldstone & Thibaut emphasize that different features can be extracted from the same stimulus object, depending on the categorization, the history of the perceiver, and pragmatic demands of the category task. This commentary addresses the influence of learning and task constraints on categorization by examining the feature strategies of novices and experts. Although novices and experts may not differ in the way they parse the parts of a domain-specific object, their featural interpretations of those parts may vary significantly.

Consider the following situation: while hiking in the woods, a novice and an expert bird watcher spot a feathered creature on the side of a tree. The novice identifies the feathered object as a "woodpecker," whereas the expert birder identifies the same object more specifically as a "hairy woodpecker." For the novice, the target category "woodpecker" need only be distinguished from other competing, genus-level categories (e.g., sparrow, hawk). On the other hand, bird expertise demands that birds be categorized more specifically at the species level (e.g., "downy woodpecker," "red-bellied woodpecker"). Thus, the pragmatics of bird expertise force the bird expert to develop more specific contrast categories than the novice. If experts and novices organize their contrast categories differently, the two groups should also differ in their functional features (i.e., features that promote within-category similarity and between-category distinctions). The novice seeks to find the functional features that distinguish between birds at the genus level, whereas the expert is sensitive to features that discriminate between woodpeckers at the species level.

How does one test experimentally the effects of learning and experience on the extraction of functional features? Schyns and Murphy (1994) used a part segmentation task in which subjects were asked to decompose categorized objects into their constituent parts. The main finding was that subjects segmented objects into different parts depending on their categorization histories. Although there is usually a good correspondence between the perceived parts of an object and their category diagnosticity in laboratory studies, the relationship between parts and features is not always so clearcut in real-world categorization.

For example, the natural shape properties (e.g., points of discontinuity, local minima) of a bird might lead both the expert and novice to parse the object into the identical constituent elements (e.g., head, legs, and beak). Although the expert and novice might agree on the constituent parts that compose an object, they do not agree on the features that describe those parts. Previous research (Tanaka & Taylor 1991) has shown that although experts list the same number of parts for objects in their domain of expertise as novices, they describe those parts in more differentiated terms or what has been called "modified parts" (Tversky & Hemenway 1984). For example, whereas a novice might describe the beak of a hairy woodpecker as "pointed" to distinguish this part from the beak of a sparrow and robin, the expert will describe the same part as "elongated and pointed." Implicit in the expert's description is information about how a hairy woodpecker's beak differs from the beaks of its more specific contrast categories of woodpeckers (e.g., downy woodpeckers). Thus, identical parts may have different feature representations, depending on the contrast categories used.

Face recognition provides another example of how the perceptual parts of an object do not correspond to its functional features. The practical demands of everyday face recognition require that faces be categorized at very specific levels of abstraction. Indeed, face recognition involves the most specific kind of categorization – *unique identity* – in which the category label (i.e., proper name) designates an individual face. Because most people can identify faces at the level of unique identity quickly and accurately, it has been suggested that virtually all humans qualify as face experts (Carey 1992).

What are the functional features that distinguish faces at the level of unique identity? According to some fixed-feature theories, such as Biederman's Recognition-by-Components model (1987), all faces are equivalent with respect to their part and configural properties (i.e., all faces contain two eyes located above the nose, which is found above the mouth). Discrimination between faces at the unique identity level must therefore be based on a fine-grained analysis of part and configural information. There is evidence, however, that this analysis does not entail the decomposition of the face into a set of independent diagnostic parts. For example, Tanaka and Sengco (in press) have shown that disrupting the configural information of one face part disrupts the recognition of another face part. This evidence supports the holistic approach to face recognition in which the functional features are not the individual face parts, but the whole face that emerges from those parts.

As the foregoing examples illustrate, the ability to quickly categorize objects at specific levels of abstraction is a trait that distinguishes an expert from a novice. Shifts in categorization level suggest that the expert is using functional features that are different from those of the novice to mark within-category similarities and between-category distinctions. Thus, the expertise work is consistent with the proposal by Schyns et al. that the featural properties of stimuli are not fixed, but are malleable to accommodate the goals of the categorization task. The study of expertise also extends the notion of a feature beyond the specification of a perceptual part. In the case of the bird expertise, the same part provides different featural information for the expert and the novice. In the case of face expertise, the many parts of a face may compose only one functional feature. Together, these studies indicate that one should be cautious when attempting to equate the perceptual parts of an object with their functional features.

Do features arise out of nothing?

Adriaan Tijsseling

Cognitive Sciences Centre, University of Southampton, Southampton SO17 1BJ, England, United Kingdom. **at@cogito.psy.soton.ac.uk**
www.soton.ac.uk/~coglab/coglab.html

Abstract: This commentary questions the validity of the claim that new features can be constructed out of nothing during categorization. A minimal set of fixed features based on what human beings are able to detect is sufficient for categorization.

Objects have features. All physically distinct objects in the world can in principle be distinguished from other objects on the basis of their features. Successful categorization depends on finding reliable invariant features so that the members of any category can be distinguished from nonmembers in a context of confusable alternatives. There are two main feature-based theories of categorization. The difference between them hinges on the answer to the question, "Where do those features come from?" According to the first theory (which we will call feature creation theory, or FCT), as proposed by Schyns, Goldstone & Thibaut, features are created under the influence of higher-level cognitive processes. According to the second theory (fixed feature theory, or FFT), we already have a repertoire of fixed features that is provided by our perceptual systems (Biederman 1987; Bruner et al. 1956; Gibson 1991; Palmer 1977). Which theory is correct depends entirely on the validity of the claim that new perceptually constrained features can be constructed, a claim I will refute here.

What people can perceive is limited. For example, we are unable to see light at frequencies lower than 360 nm or higher than 760 nm; we cannot hear sounds of frequencies above 20,000 Hz (Coren & Ward 1989). The human perceptual system has a specific architecture and function, both of which have evolved over millions of years. These systems are equipped with specific detectors and neural wiring (Barlow 1979; Fujita et al. 1992; Gross et al. 1972; Hubel & Wiesel 1965; Kobatake & Tanaka 1994; Livingstone & Hubel 1988; Maffei 1978; Salzman et al. 1990; Snippe & Koenderink 1992). What we can detect determines our minimal basic set of primitive features (e.g., line segments, angles, colors). Every other feature is composed of the fixed set of primitives. Feature construction is guided by repeated presentation of a stimulus. For example, an "A" encountered for the very first time may just look like a three-line scribble, but with every other presentation, the letter A becomes a feature itself and as such it will become much easier to detect in a page full of scribbles.

I am not of course claiming that "responding to a feature" – in the sense of a biophysical difference in receptor activity with one stimulus and another – is the same as detecting the feature. It is possible that finding the invariant feature among a chorus of low-level detectors is where the real cognitive work begins. We need to select – from all the features to which we are sensitive – the right ones for a particular categorization task. Our brain cannot create features out of *nothing:* it must be selected from the sensory elements of which vision, hearing, and touch are composed. Our perceptual systems evolved to detect what needs to be detected, that is, nature gives us the primitives. The action may not be at this level of primitives, but that is the reason our cognitive and perceptual abilities include the power to combine and recombine (but not to create) features.

If there is a set of primitive features, what does it look like? According to Schyns et al., a set of primitives that could ground categorization must satisfy at least three conditions: (1) primitives must exist before experience with the objects they represent, (2) they must be sufficient to represent the entire set of representable objects, and (3) they must be able to bootstrap complex recognition systems. If primitive features exist before perception starts, then they must in some sense be present in the brain from birth. If they must further be able to represent what can be represented, then the only way to find those primitives is to look at what the perceptual systems do with sensory information. I have already noted that the perceptional range is constrained (unaided by instruments), and that our perceptual systems evolved to embody specific detectors. If the function of one of these detectors is to fire when a vertical line is present, then a vertical line is presumably one of the primitives.

How primitive can a primitive be? Schyns and Rodet (1996) argue that fine-grained primitives such as pixels cannot be sufficient for categorization even with powerful rules of combination. I agree that the presence of such a low-level primitive as a pixel would make little sense, given the existence of a vertical line detector. Nature has already discovered that a two-dimensional proximal image on the retina is best analyzed using colors, line segments, and curvature. There could of course be a primitive for the presence of one black dot, but the perceptual system is likely to use its more complex primitives as often as possible, because they allow quicker responses and analyses. The idea that primitives can be complex features such as curves, corners, or edges has been proposed by, among others, Garner (1974) and Biederman (1987). An argument against their view is that recognition systems that use complex primitives are blind to structures that are not represented as primitives and are not composed of simpler primitives (Schyns et al.). This criticism only holds if the set of primitives is not enough to represent the entire set of representable objects and if it does not include less complex primitives such as small black dots.

The process of feature detection itself is crucial here. One can ask how the features in a retinal image are to be found (Yulle & Ullman 1990). The idea is that the system first checks whether its more complex primitives are present in the image and only resorts to more basic and finer grained primitives if this fails. The system is trying in both cases to construct or combine new functional features to speed up recognition in future encounters with similar objects. So although a new functional feature would be composed of several primitive ones, and hence would not increase the system's representational capacity, it could still increase its efficiency. Having the new feature "A" might not allow us to represent more, because it is composed of more primitive line features such as "/," "-," and "\," but it does increase the speed with which we can recognize the letter in incoming sensory information.

To illustrate that a fixed feature set is not so rigid, one can argue that the experiment of Schyns and Rodet (1997) described in the Schyns et al. target article as supporting FCT, can be explained using FFT. The reasoning here was that subjects should categorize X-Y cells as XY members if they perceive and represent XY as a conjunction of two individuated features. It was shown that the $X \rightarrow Y \rightarrow XY$ group performed this categorization and that the other group classified X-Y as either X or Y. Schyns and Rodet (1997) suggest that this result challenges FFT's main claim that category learning consists only of weighting the features of a fixed set. However, this argument holds only if FFT claims that a fixed set of features really contains *all* the possible features present; it does not hold if a fixed feature set means a minimal fixed set of primitives out of which all other features can be composed. The features x, y, and xy can be composed of a minimal primitive set, and the order in which the features are constructed and their categories are learned could lead to an orthogonal classification of X-Y. Hence, this experiment does refute "strong" FFT, but it does not seem to rule out "weak" FFT.

Is it absolutely impossible to create features? No: there is and should be a way out. Evolution takes care of cases in which the environment might change in such a way that some primitive features lose their utility and other primitives are needed. For example, suppose the sun changed in such a way that light consisted only of wavelengths above 500 nm. The color blue would no longer exist and the "blue" receptor cone would lose its function. Future generations might then evolve a new version of these receptors that is responsive to a different kind of light.

Who needs created features?

Katja Wiemer-Hastings and Arthur C. Graesser

Department of Psychology, The University of Memphis, Memphis, TN 38152-6400. **kwiemer@cc.memphis.edu; a-graesser@memphis.edu www.psyc.memphis.edu/faculty/graesser/graesser.htm**

Abstract: Schyns, Goldstone & Thibaut present reasonable arguments for feature creation in category learning. We argue, however, that they do not provide unequivocal evidence either for the necessity or for the occurrence of feature creation. In an effort to sharpen the debate, we take the stand that a fixed feature approach is to be preferred in the absence of compelling evidence.

The notion that categorization requirements influence perception is not a new one. A few decades ago, Whorf proposed that the language in a culture creates categories that guide our perception of the world (see Hunt & Agnol 1991, for a discussion). If this is true, then the question arises: What forms of learning are used in this process?

The findings reported by Schyns, Goldstone & Thibaut do not directly resolve the issue of whether new features are created during learning. Instead, they could be explained by a finite set of fixed features that are hardwired in different modular systems (e.g., particular sensory modalities, spatiality, motor skills, language, deeper components of cognition). Through experience, these fixed sets of features are noticed, enacted, differentiated, weighted, and combined through chunking (Anderson 1990; Newell & Rosenbloom 1981). Moreover, the fixed-feature approach offers more tractable models that are grounded in the theories associated with each modular system.

Schyns et al. argue that an individual's history of categorization produces different sets of features. In the experiment by Schyns and Rodet (1997), two groups learned categories in a different order and were shown to extract different feature sets. An alternative account of the finding is that a fixed set of features was weighted differently by the two groups. Furthermore, it does not follow from their results that different categorization histories *permanently* produce different feature sets. Instead, it is conceivable that if the same categories were presented to both groups over and over again, both groups would (by means of differentiation and feature weighting) end up with the same feature set. Evidence of feature creation would be stronger if it could be shown that the groups would keep different feature sets even after repeated exposure to the same categories and exemplars.

Schyns et al. correctly point out that the traditional experimental material in concept learning experiments biases the results toward finding fixed feature sets. However, the alternative material suggested by the authors might produce a bias in the opposite direction. That is, in the categories that prevail in "everyday life," the requisite site feature discriminations are rarely as fine as those required in their proposed material.

Categories outside expert domains are typically vague with respect to the membership of some exemplars. This vagueness is partially explained by the resolution level of the salient features at this "novice" level: they are less precise than those used for expert categories. Malt (1995) discusses studies that suggest that "discriminations finer than those required for categorization at the generic level, although possible, are not salient in the absence of greater than average attention" (1995, p. 124). To make finer, more precise distinctions and subcategories (i.e., in expert domains), the attention window has to be shifted or zoomed in to more precise features; these features already exist and are automatically being used by the perceptual system. These fine-grained features are also processed through vague categories, but they are not attended to because they are processed in an automatized fashion. The difference between the categories of the novice and the precise categories of the expert is compatible with the view that humans ordinarily operate on the basis of a limited, economic set of fixed features. We believe that the stimulus material proposed by Schyns et al. is representative of categorization in expert domains.

Malt has a compelling discussion of anthropological studies that examine how much categorization is influenced by cognition versus utility in a culture versus the environmental structure. She argues that expert categorization requires an attentional shift to a different level of feature abstraction because very similar objects have to be distinguished. Finer-grained distinctive features become salient in this process and receive more weight. For example, categorization of Martian cells requires discrimination at a very precise feature resolution. Schyns et al. argue that category learning in this expert domain requires feature creation. In contrast, we argue that expert category learning could still be based on a set of fixed features. These features are theoretically grounded by the constraints of the world, sensory organs, neurophysiology, evolution, and higher-level cognition.

There are at least two ways in which fixed features can suffice for the acquisition of categories. First, a set of fine low-level features are chunked to flexibly accommodate the new categories. Second, the relevant features that are discriminating between categories are represented as "fuzzy" high-level features that are correlated with a subset of the fixed features. Hence, we are not convinced that feature creation is a necessary process. Instead, fixed features may be sufficient to represent a wide range of object categories, for both novices and experts.

Feature learning during the acquisition of perceptual expertise

Pepper Williams,[a] Isabel Gauthier,[b] and Michael J. Tarr[c]
[a]*Department of Psychology, University of Massachusetts Boston, Boston, MA 02125.* **pepper@umbsky.cc.umb.edu**
psych.umb.edu/faculty/williams/williams.htm
[b]*Department of Psychology, Yale University, New Haven, CT 06520.*
isabel.gauthier@yale.edu
www.cog.brown.edu/brochure/people/mgt/peoplegauthier.html
[c]*Department of Cognitive & Linguistic Sciences, Brown University, Providence, RI 02912.*
michael_tarr@brown.edu www.cog.brown.edu/tarr.html

Abstract: Does feature evolution stop once we have acquired sufficient features to perform a recognition task? With extended practice, novices may develop a more sophisticated feature space that allows them to perform more accurately or quickly. Our work on perceptual expertise indicates that feature learning and reorganization can continue even after an initial set of features is available to represent a novel class of objects.

Schyns, Goldstone & Thibaut argue that new features will evolve when an object class cannot be represented using previously developed features. We are sympathetic to the authors' point of view that the human object recognition/categorization system is plastic, without a fixed feature vocabulary, but does feature evolution necessarily stop once we have acquired sufficient features to perform a given recognition task? A novice birdwatcher may quickly develop a collection of features for distinguishing different species of hawks, but this feature set may not be ideal. With extended practice, novices may be able to develop a more sophisticated feature space that allows them to perform more accurately and/or quickly. Our work on perceptual expertise (Gauthier & Tarr 1997; Gauthier et al. 1997b) provides support for the idea that feature learning and reorganization can continue even after an initial set of features is available to represent a novel class of objects.

The stimuli we have used, "Greebles" (Fig. 1a), are easily decomposable into constituent parts. Moreover, participants unfamiliar with Greebles (novices) can learn to identify individuals without difficulty, indicating that people in our participant pool (undergraduates) either already have the features necessary to categorize Greebles, or can develop the requisite features almost immediately. According to Schyns et al., these conditions should lead to "fixed-space" learning: distinctions between different Greebles should continue to be made using the features participants use during their initial encounters with the objects.

However, when participants were trained for many hours on Greeble recognition (Gauthier et al. 1997b), we found that their response times on a Greeble-naming task continued to go down even after they reached near-perfect accuracy levels (Fig. 1b). Furthermore, correlational analyses of the response time data showed that the Greebles participants found easiest to recognize at the beginning of training (when they were novices) were not necessarily the easiest to recognize once they became experts. These findings indicate that perceptual learning, and possibly feature differentiation, continues even when features sufficient to recognize the Greebles have already been acquired.

Once the training regimen was completed, these "Greeble experts" learned to name new Greebles faster than novices did, and more important, showed qualitative differences, compared with novices, on tests such as the Tanaka and Sengco (in press) old/new configuration task. In this test, participants are asked to identify one portion of a known Greeble that is presented either in the normal (old) Greeble part-configuration or in a transformed (new) configuration, for example, with the top, side-attached parts rotated 15 degrees around the vertical axis toward the front. Experts, but not novices, were significantly impaired at recognizing parts in new configurations (Gauthier & Tarr 1997; Gauthier et al. 1997b), again indicating that experts had developed qualitatively different ways of representing Greebles even though "novice features" could have provided a sufficient basis for Greeble identification.

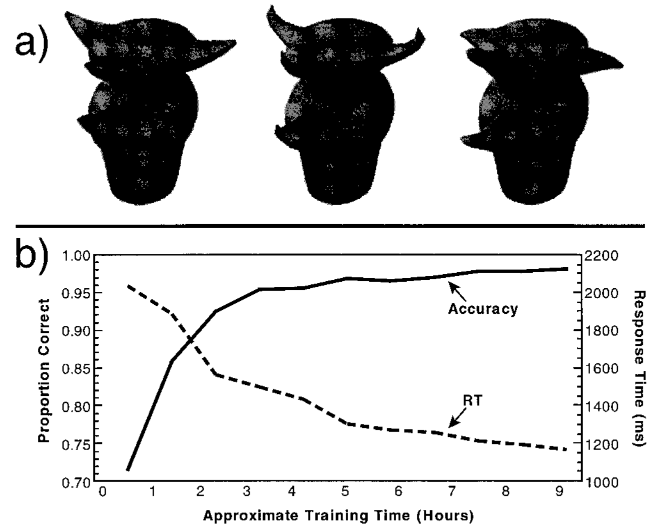

Figure 1 (Williams et al.). (a) Three Greebles from our studies of perceptual expertise. All Greebles share the same part structure, but each one has uniquely shaped appendage parts. (b) Training results from a Greeble expertise experiment. The plot shows accuracy and response times for a series of tests (performed over the course of 10 one-hour training sessions) in which participants had to name up to 20 individual Greebles (from Gauthier et al. 1997b).

These and other results from our studies suggest that the simple featural contrasts that may be used by perceivers when they first learn to discriminate between members of an object class may not be used by the same perceivers once they become highly familiar with the class. Although broadly consistent with the feature-creation framework of Schyns et al., our findings challenge the proposition that the feature space ever becomes fixed. In other words, an expert's feature space may become reorganized in response to environmental pressures to perform a categorization task more efficiently. Perhaps *every* encounter with an object of a class leads to a small amount of feature-space reorganization. Such a mechanism would not only lead to constant improvements in performance (as long as such improvements are possible), but it would also do away with the need to "decide" when a given learning task requires fixed-space and when flexible-space learning.

If our hypothesis is correct, then at least two important questions need to be answered. First, what are the environmental pressures that cause an expert's feature space to become reorganized? In our studies, participants were explicitly instructed to perform as quickly as possible; similarly, our birdwatcher would be under similar time pressure, as he may get only a fleeting glimpse at a to-be-identified hawk. A natural history museum curator, on the other hand, would have ample time to examine birds for his exhibits, but would want to be exceedingly accurate. Yet another form of expertise might be exhibited by a falconer, who needs to identify the best way to handle individual hawks. We hypothesize that these situations would all lead to different feature spaces, even though the visual stimuli would be the same in each case.

A second issue is how feature reorganization might be accomplished. Preliminary data from a longitudinal fMRI (functional magnetic resonance interferometry) study (Gauthier et al. 1997a) indicates that a particular area in ventral temporal cortex may become increasingly important over the course of training for Greeble recognition. Neurons in this area thus appear to be particularly well adapted for processing and encoding features that support the fine metric discriminations needed for fast and accurate identification at the individual level, but extensive training seems to be necessary to tune these neurons to the particular types of features found in a given object class.

Authors' Response

Ways of featuring in object categorization

Philippe G. Schyns,[a] Robert L. Goldstone,[b] and Jean-Pierre Thibaut[c]

[a]*Department of Psychology, University of Glasgow, Glasgow G12 8QB, Scotland;* [b]*Department of Psychology, Indiana University, Bloomington, IN 47405;* [c]*Department of Psychology, Université de Liège, Sart-Tilman 4000 Liège, Belgium.*

philippe@psy.gla.ac.uk www.gla.ac.uk/acad/psychology; rgoldsto@ucs.indiana.edu. cognitrn.psych.indiana.edu/; thibaut@ulg.ac.be

Abstract: The origin of features from nonfeatural information is a problem that should concern all theories of object categorization and recognition, not just the flexible feature approach. In contrast to the idea that new features must originate from combinations of simpler fixed features, we argue that holistic features can be created from a direct imprinting on the visual medium. Furthermore, featural descriptions can emerge from processes that by themselves do not operate on feature detectors. Once acquired, features can be decomposed into component features if required by other categorizations. We therefore argue that it is not necessary to separate holistic and componential approaches to representations, because the latter is a development of the former. The requirements for representational flexibility outstrip the performance of any existing computational models, but specific mechanisms of feature creation are discussed and evaluated. Challenges for feature creation mechanisms are discussed together with the constraints (perceptual, statistical, functional, and task) they will need to satisfy.

R1. Introduction

Objects form categories because they share a number of features and differ on other features from contrasting categories. One essential function of a feature, therefore, is to subserve the categorization and representation of objects. The target article inquired into the origin of features. Either people come equipped with a complete, fixed feature repertoire that accounts for all present and future categorizations, or they sometimes create flexible features to subserve new categorizations. The target article examined the implications of the latter view: that flexible features (1) are not necessarily derivable from a fixed set of primitives, (2) can augment the representational power of the feature repertoire, and (3) can change the way an input is perceived.

R2. Componential and holistic object representations

From the outset, it is important to emphasize that the flexible feature stance adheres to the reductionist enterprise of cognitive psychology according to which complex object representations should be reducible to combinations of their building blocks. One of the major reasons why objects are described in terms of features (instead of pixels) is because of the information compression and the resulting gains in efficient encoding that features confer on the system. We do not contest a reductionist approach, as **Huettel & Lockhead** suggest; rather, we challenge the possibility that all object representations can be reduced to a fixed set of building blocks. We are compositionalists in that we assume that features, once created, become the building blocks of future object encodings and representations. However, a flexible feature repertoire must be maintained because there will always be differences between objects that a fixed repertoire will not represent. **Braisby & Franks** argue that combinations of fixed blocks could very well represent these differences. To this we reply that if the representational granularity of fixed blocks is above the granularity of the pixel (or any similarly unstructured primitive), combinations of these blocks could not represent an object difference that the representational resolution of the fixed blocks does not capture. We will argue later that pixels are not representations of forms, as features are, but are a medium for the representation of forms.

According to the main alternative to the compositional view, objects are represented holistically, without a prior decomposition into their components. **Edelman,** for example, discusses a system that learns the appearance of entire objects and uses their memory traces as holistic features. In a related vein, **Burgund & Marsolek** suggest that specific brain systems could be dedicated to the extraction of holistic features that represent familiar objects. Holistic representations have two main advantages over componential ones: (1) They preserve the input in an unprocessed form, and (2) they represent the input efficiently by compressing multiple sources of information into a single feature. Holistic and componential representations can coexist in the flexible feature framework. It is therefore an "augmented" compositionalism.

As already discussed in section 1 of the target article, compositions of flexible features retain all of the strengths of the classical "mental chemistry" of the fixed feature approach: many object descriptions can be generated from a set of elements and a set of combination rules; feature combinations allow for structured hierarchical representations. Similarity relations between different objects can be expressed in terms of their features and their combination rules.

In the target article (sect. 2.8, para. 6), we argued for the addition of feature decomposition mechanisms to supplement the classical mechanisms of feature composition. Feature decomposition is a mechanism that breaks down a holistic feature into two or more holistic subfeatures when one of them is required to represent a new subcategory (see Schyns & Murphy 1994, experiment 2). The importance of feature creation together with decomposition should not be underestimated in componential theories. Created holistic features share the two properties of holistic representations listed above: they preserve the raw properties of the input because their decomposition into subunits is contingent on the task that requires the decomposition and created holistic features are representationally compact because decomposing into subcomponents is not always required. Thus, flexible features do not impose the usual dichotomy (other than phenomenological) between holistic and componential representations. Holistic and componential representations can coexist within the scheme because the latter is a further development of the former.

In sum, our proposal for flexible features adheres to a reductionist approach to cognition: Flexible features and combination rules form the basis of object analyses, representations, and categorizations. Throughout development, new holistic features that preserve the raw properties of the input are created to represent new categorizations. Task-driven decompositions of these features give rise to subfeatures, which remain holistic themselves until another categorization requires their decomposition into sub-subfeatures. Schyns and Murphy (1994) argued that the decomposition principle could account for the emergence of conceptual hierarchies throughout development: decomposition determines the holistic-to-componential link, and chunking drives the componential-to-holistic link. We expand on chunking in the following section.

R3. What does "new feature" mean?

This section discusses different ways in which a feature can be new. First, we review the production of newness through the chunking of simpler components and then discuss how newness can also emerge from a process that directly imprints on a continuous, a priori unsegmented medium (a feature creation process). We then outline how new properties could emerge from imprinting.

R3.1. Chunking

Chunking and perceptual unitization were discussed in the target article (sect. 3.1) as processes that could synthesize new features from a set of more elementary components. Chunking therefore implies that the input is segmented into features before being chunked, in much the way a capital T may initially be discretized into a vertical and horizontal bar before being chunked into a holistic "T." Phenomenologically, the perception of the chunk does not entail the perception of its components. A chunk is therefore a new, isolable, and independent information packet of psychological processing that has the properties of holistic representations: The recognition of a chunk will not necessarily prime the recognition of its components, and vice versa. Response times to the chunk will be faster than predicted by response times to the components (Goldstone et al., in press), and similarity relationships will not necessarily be perceived between the chunk and its parts.

Good examples of chunking producing new features have recently been described in the category of expertise literature. For example, **Tanaka** discusses faces that are initially analyzed into their consituent parts (e.g., nose, mouth, eyes, and ears) before being chunked into configural, holistic representations with the progressive acquisition of expertise in face categorization. Similarly, **Williams et al.** propose that extensive experience with objects designed to be geometrically as homogeneous as faces (the "Greebles") induces the creation of new features. In the discussed experiments, tests on the nature of the features gave support for holistic representations. These results suggest that extensive experience with objects that look alike initially can, by chunking already existing features, induce a synthesis of new configural encodings that increases the contrast between the objects. The main property of these features is that they are initially derived from, but ultimately do not reduce to, component parts.

R3.2. Feature creation: Form-from-medium rather than form-from-form

We wish to distinguish the form of feature newness involved in chunking (which we call form-from-form) from another kind (form-from-medium), which produces features directly from a medium, not from other features. By analogy, imagine a Martian whose visual medium (the output of its transducers) is very much like dough. On the first day of its existence, the outside world imprints a teddy bear into the dough. Of course, this object and its parts are unknown to the Martian, who cannot compose a new representation of the modelled teddy bear from an already-existing representation of its component parts. Feature creation is a process that can make a direct imprint on the visual medium: it can "cut around" the teddy bear's silhouette and represent the entire object as a new holistic feature. Note that this cutting process only separates the entire bear from the medium. At this stage of conceptual development, our Martian represents the bear as a unitary, holistic feature and its decomposition into subcomponents is unspecified.

The dough analogy illustrates that forms (i.e., features) can arise from the absence of form (i.e., a medium) with proper "cutting principles" (i.e., generic perceptual constraints). Although form-from-form dominates compositional approaches (see **Dominey** and **Tijsseling**), we believe that form can also arise from a high-dimensional medium and adequate perceptual constraints. We must start with a rejection of the idea that pixels, or retinal outputs are already fixed features (as suggested by **Dorffner** and **Dominey**). We believe that the proper relationship between pixels and features should mirror the relationship between dough and form: the former is the medium for the expression of the latter. In other words, individual pixels do not represent forms, but together, millions of pixels serve as a high-dimensional medium for multiple expressions of form. For example, **Edelman**'s system can make a direct imprint on pixels, which serve as a medium. However, perceptually constrained imprinting will already involve a massive reduction of dimensionality from the medium (see sect. 3.4.2 of the target article). This is developed further in the next section.

R3.3. Emergence of new properties

Hummel & Kellman ask how new abstract properties (e.g., squareness) could ever emerge from feature creation if they are not derivable from the properties of lower level features. Although this is a particularly difficult issue, a partial answer can be developed by considering perceptual constraints and the imprinting process. We argued earlier that chunking could produce configural features (e.g., face configurations, see also **Tanaka; Williams et al.**) that do not reduce to the simple additions of their components (e.g., face parts). This is an example of the nonlinear recoding that Hummel & Kellman discuss, but they quite rightly separate emergent abstract properties from nonlinear recoding.

A less obvious form of property emergence could arise from the massive dimensionality reduction that accompanies imprinting. Remember that imprinting extracts features from the visual medium directly in response to task constraints. We also argued in the target article (see 3.4.2.), however, that perceptual constraints should play a critical part in this reduction. Properly constrained dynamical systems can have interesting emergent behaviors that are similar to those we have in mind. For example, Shashua and Ullman (1988) describe a process that is tuned to extract shapes with silhouettes that satisfy constraints as to their length and smoothness. The extraction is a dynamic relaxation process that preserves only smooth and long contours while erasing discontinuous and short contours (as noisy contours tend to be). Although the system has no detectors for specific silhouettes (i.e., no fixed features for these shapes), its internal constraints implicitly define a gradient of "perceptual" sensitivity to different silhouette classes. For example, and to simplify a little, the system would be more sensitive to a square than to a circle of the same area if its constraints favored long straight contours. Ullman (1984) also provides examples of simple visual processes that can act like detectors for specific object properties (e.g., their closedness), without directly representing these properties in the system. By analogy to these processes, the dynamics of dimensionality reduction could be biased to first produce shapes that are closed, have good continuation, are smooth, or satisfy the constraints described in **Singh & Landau, Quinn, MacDorman** and **Benson.** Shapes that would fit well with the system's perceptual constraints (e.g., symmetric, closed shapes with simple geometries) would be perceptually more salient, relatively easier to extract from the medium to form new features, and more likely to give rise to emergent, nonderivable perceptions.

In sum, we believe that the emergence of nonderivable properties could be grounded (at least in part) on the perceptual constraints that guide the dynamical process of dimensionality reduction. Although this does not preempt the question of property emergence, it might provide a first step toward its understanding.

In the context of self-organizing machines, Cariani (1993) discusses a very interesting example of forms arising from interactions of external signals, a medium, and specific constraints: the Pask device (Pask 1960). The Pask device was designed in the 1950s as a system that could create its own primitive features. It consists of an array of platinum electrodes, receiving signals from the outside world, that are partially immersed in an aqueous solution of metallic salts (e.g., ferrous sulfate). This unstructured medium has the capacity to grow new structures. Passing current through the electrodes literally grows dendritic metallic threads between them. The growth of the dendritic structure can be controlled by choosing the electrodes between which the current will pass. It is worth noting that this aqueous solution has no a priori structure, but new structures having emergent properties could be created in response to external inputs. Such a device could physically implement the computations of a Perceptron in an analog medium – the conductances between electrodes in the array would correspond to the connection strengths of a perceptron (Cariani 1993). However, because the dendritic tree itself could evolve, the system could change its entire computing architecture – that is, not just its weights, but also its pattern of physical connections. In fact, Cariani (1993) reports that within half a day, ferrous threads could be grown to become sensitive to a sound or a magnetic field. With more time, the device could discriminate 2 frequencies (50 vs. 100 cycles per second).

This example brings together the ideas of form-from-medium and of emergent properties from a system's dynamics, as discussed earlier. Given a medium and specific constraints (here, ferrous sulfate, electricity, and the growth of dendritic structures), new forms (dendritic structures) can evolve that confer on the system emergent properties (the categorization of sounds) that are not reducible to preexisting properties. Pask's (Pask 1960; Carlani 1993) device does not originally have frequency detectors, only processes that can build other processes that detect frequency differences.

R3.4. What is a feature like?

The discussion so far has shown that chunking, imprinting, and emergent perceptual properties are different forms of feature newness. These distinctions made implicit assumptions about that nature of features that we now highlight.

Features are isolable information packets. As a result of this first property, features represent compressed information from the input to which the system is sensitive. This information can be parts (as suggested by **Tanaka** and **Singh & Landau**), but it is not limited to parts. Colored blobs, textual elements, and many other information packets can qualify as features as long as the system's psychological response to the packet reveals that it is a discrete, holistic entity in psychological processing.

Features are independently perceived. Unlike feature conjunctions, isolable features are perceived without the perception of each component preceding perception of the conjunct. Independently perceived features can evolve from a chunking process that produces configural features, from an imprinting process that creates a new form from the visual medium, or from dimensionality reduction itself.

Behavioral measurements can be gathered to assess the perceptual independence of features with respect to their components (e.g., Schyns & Rodet 1997). Behavioral data are better than the type of logical analysis advocated by **Latimer** at distinguishing the wholes from the parts in psychological processing. One difficulty with logical analysis is that it does not address the issue of the perceptual independence of the whole with respect to its parts. This is particularly important in our framework because, as already discussed in the context of feature decomposition, a holistic feature at time t of conceptual development might not have specified components. These will only be determined at time $t + 1$, when a new categorization forces a specific decomposition. The indeterminacy of parts with respect to their wholes might even preclude a logical analysis that does not integrate the individual's history of categorization.

Features provide a perceptual organization of the input. Isolable features provide a stable analysis of the ever-changing sensory output for subsequent processing such as similarity assessment, categorization, reasoning, and adaptive action. In this sense, isolable features perceptually organize the input (see also **Singh & Landau**). This featural organization bridges the gap between the physical world and cognition.

R4. How much flexibility?

Commentators challenged our view of flexibility in two opposite ways. Some thought that our flexible features are too flexible (**French & Weaver; Palmeri; Tijsseling; Wiemer-Hastings & Graesser**) because a system can do most or all of its tasks with preexisting features. Other commentators agreed that learning systems would at least need the flexibility we discussed, and perhaps even more (**Brooks; Deák; Hahn & Chater; Quinn; Tanaka; Williams et al.**).

R4.1. Does feature creation require more flexibility?

R4.1.1. Continuous versus discrete feature learning in time. One possible interpretation of our proposal is that features created to solve a categorization subsequently crystallize and become fixed. Several commentators objected to this, saying that we should instill more flexibility in the feature repertoire (**Hahn & Chater; Tanaka; Williams et al.**) because feature creation and adaptation is a never-ending process. We agree that the feature repertoire should evolve continuously in response to the flux of new categorizations facing the organism. One could even suggest that every encounter with the objects of a class could lead to a small amount of feature-space reorganization (Williams et al.; Hahn & Chater). The issue of continuous variability raises an important problem that was not discussed in the target article: How does the system store information about feature variations? One possibility is that each time the system sees an object, it replaces its old feature representation with new features. However, such a system is susceptible to catastrophic forgetting. Successive exposures to rare instances of a feature could lead to a transformation of the feature that would no longer match common instances. Another possibility for representing variations is to apply template matching to features and store each transformation as a distinct feature exemplar in memory. However, given all the potential variations in illumination conditions, position, and occlusion that features can undergo, exemplar-based systems would rapidly be overwhelmed with the millions of feature exemplars they would need to store and organize. We therefore favor an abstract representation (for parts, for example, think of a parametrized elastic template) that would adjust very slowly to local feature transformations to preserve a desirable inertia in the continuous adaptation to changes. The next section discusses how this might occur.

R4.1.2. Feature transformations in the context of a category

Feature transformations introduce the idea that the system should tailor its features to the contingencies of the learning episode (**Hahn & Chater; Lamberts; Williams et al.**). Hahn & Chater suggest that we should frame our features in a transformation space because different categorizations of an object set never imply the same subset of features. Pushing the argument forward, stable features might not be necessary at all.

We agree that transformations of a feature within a category are useful for learning the possible variations of the created feature. Schyns and Murphy (1994, experiment 4) report an experiment in which subjects were shown one category of 2D objects composed of two parts. One part transformed randomly across instances and the other transformed systematically. The systematic variations were varying degrees of protrusion and extrusion of a blob of the part. In such an experiment, subjects must create part features to represent the category. Because only one of the two parts varied systematically, subjects could create a feature only for that part. This is of course possible to the extent that the transformations of the part across instances do not make it appear to be different parts, rather than different transformations of the same part. The outcome was that subjects represented part transformations with two distinct parts: one for all the extrusions, one for the protrusions. Thus, even though the context of the category served to represent the part and its transformations, there was a strong perceptual limitation on the transformations a single part could "tolerate" and still be judged similar: parts had to share the same qualitative protrusions and extrusions (technically, signs of curvature) across instances.

Not all encodings of feature transformations need to be permanent, however. **Lamberts** suggests that there are transient changes in specific context. Repetitions of transient changes should be a good indicator that they become more permanent. This is reminiscent of a neural network system with transient weights that can adjust rapidly to changing contingencies while long-term weights maintain the integrity of the computing architecture. Such mechanisms are compatible with the idea of flexible but stable features.

R4.2. Feature creation: Pervasive principle or limited applications?

One argument against flexible feature creation is that the hypothesized principles were only demonstrated in very restricted situations of categorization. The materials that show evidence of feature creation are qualitatively different from the objects people typically categorize. Hence, the fine discriminations these materials require go far beyond those required in everyday recognition (**Wiemer-Hastings & Graesser**). Similarly, **French & Weaver** argue that the feature creation we advocate is limited to the acquisition of expert categories. Therefore, we argue for more flexibility than is really necessary for everyday categorizations.

On the contrary, we believe that everyday categorizations require very subtle discriminations. Instances of these discriminations occur when people identify faces, distinguish a Jonagold apple from a Cox, differentiate a poodle from a basset hound, or recognize their own car amongst many others. These cases of everyday recognition involve objects that require discriminations at a very high resolution. Of course, these are also fine-grained categorizations, but there is no principled reason why feature creation should be limited to the "subordinate world," as implied by **French & Weaver** and **Williams et al.** We believe that feature learning mechanisms can also account for the development of basic-level and superordinate categories (insofar as the latter requires perceptual encodings). If the basic level is the most inclusive one at which objects look alike, the entry point to recognition could change with the acquisition of new features (see Schyns, in press, for a discussion). To illustrate, face identification is often thought to be a clear-cut subordinate categorization (because all faces have a similar global shape). However, different views of the same person look more alike than the same view of different persons. Thus, our basic categorization could be at the level of the individual (the level at which face views are more alike) instead of their assumed "basic" level, the face categorization (the level at which face views look more different). However, this effect could be inverted with the faces of another race, for which discriminative features remain to be acquired. Faces that are generally less familiar might look more similar from the same viewpoint. Hence, the basic level of a category could very much depend on the features that optimize the perceptual distinctiveness of this category (parts, configurations, color, texture, and so forth). We saw earlier how as one achieves expertise nonlinear encodings of parts could enhance the distinctiveness of an object set (**Tanaka, Williams et al.**). Thus, the subordinate level category for novices might be the basic level category for experts and the perceptual appearance of identical objects could change dramatically. If expertise for a given category changes across individuals, it could have profound implications for theories of object recognition and categorization. As Tanaka and Taylor (1991) elegantly put it: the basic level could be in the eye of the beholder.

In sum, future research may reveal that feature creation applies only to the learning of very specialized (or subordinate) categories such as X-rays, dermatosis, and so on. Alternatively, it may turn out that feature creation mechanisms pervade the very early stages of conceptual development (when the first categories and their structuring features must be learned), but that mature categorizers (who tend to know the relevant perceptual analysis of most objects) create new features only when they learn expert categorizations. Whether feature creation mechanisms have a broad or a limited impact on the development of familiar object concepts is now an empirical issue of developmental psychology.

R4.3. The loci of plasticity

It is argued by **Tijsseling** that elementary features are fixed, and that flexibility derives from novel combinations of these elementary features. Neuropsychological evidence does not provide support for this position in that neural plasticity is found at very early stages of sensory processing. For example, practice in discriminating small motions in different directions significantly alters brain electrical potentials that occur within 100 milliseconds of the stimulus onset (Fahle & Morgan 1996). These electrical changes are centered over the primary visual cortex, suggesting plasticity in early visual processing. Karni and Sagi (1991) find evidence, based on the specificity of training to eye (interocular transfer does not occur) and retinal location, that is consistent with early, primary visual cortical adaptation in simple discrimination tasks. The above evidence suggests that perceptual plasticity may arise even earlier than the visual area V4 proposed by **Benson,** although V4 is probably also a site of plasticity and the above claims will require empirical confirmation. In audition, classical conditioning leads to shifts of neural receptive fields in primary auditory cortex toward the frequency of the rewarded tone (Weinberger 1993). **Burgund & Marsolek** suggest that the right hemisphere could participate in the abstraction of holistic stimulus properties, whereas **Williams et al.** stress the potential relevance of an area in the ventral temporal cortex during the course of expert learning. In short, there is an impressive amount of converging evidence that experimental training leads to changes in very early stages of perceptual processing. These results are also interesting in that they all occur in mature animals. The creation of hypercolumns in the brain throughout development has been known for a long time (Hubel & Wiesel 1977) and implemented in a multitude of neural networks (e.g., von der Marlsburg 1973). We agree with Benson's argument for a particularly robust period of perceptual plasticity in early development. However, this stage is not a "critical period" in that plasticity in perceptual systems continues, if at a decelerated pace, well into adulthood.

R5. Mechanisms of feature development

Several commentators (e.g., **Gauker, Hummel & Kellman, Palmeri, Phillips**) took us to task for not proposing a particular mechanism for the creation of new features. Our primary objective in the target article was to describe an overlooked but necessary process of constructing featural representations that are required for many category learning tasks. It was not our purpose to propose a specific implementation of this process. In fact, we suspect that the theoretical analyses of the requirements for representational flexibility and the constraints on feature learning far outstrip the performance of any existing computational model. Nevertheless, if no mechanisms exist that could conceivably generate novel features, then our approach would be open to criticism as nonimplementable, at least given our current understanding.

R5.1. Specific proposals

Fortunately, many computational models have implemented aspects of our proposals. The commentators describe several of these models, many of which are neural networks. Neural networks are well suited for our proposed requirements because they typically have intermediate "hidden" units that intervene between inputs and outputs and can be interpreted as the system's acquired features (**Dorffner, Hummel & Kellman**). Several aspects of the positive proposals for feature creation are worth noting. **Abdi et al.** propose that the statistical procedure of Principle Component Analysis (PCA), which has been implemented by neural networks, can generate the dimensions of a set of faces by determining the largest components of variation in a set of pixel-based faces. They also propose supplementing this process of dimension generation with task information that selectively weighs regions of pixel space that are important for categorization. PCA has the desirable property of compressing a large amount of information into a single component (which Abdi et al. construe as a holistic feature).

The proposal for feature generation via information compression is supported by neurophysiological evidence suggesting massive information reduction from primary sensory maps to later neural stages (**Phillips**). **MacDorman** suggests that a simplistic interpretation of this neural process as computing PCA is inappropriate. If neurally implemented, the entire PCA procedure would need to be recomputed after the presentation of each object. Fortunately, neural network models exist that can incrementally adapt the space on which objects are described without requiring the reanalysis of all previous objects, and without requiring the creation of orthogonal components.

In addition to proposals for creating new features from compressed representations, a second common proposal from the commentators was to develop units specialized for the processing of specific instances. That is, units would be formed to represent entire stimuli (**Burgund & Marsolek, Edelman, Tanaka, Williams et al.**). As pointed out by Burgund & Marsolek, in many cases if processing entire stimuli becomes more highly efficient over time it is probably the result of a process of whole-object imprinting. These proposals gain support from FMRI studies finding that areas in the ventral temporal cortex become specialized for complex configurations (Williams et al.).

In short, there has been significant progress on both computational and neurological mechanisms that could implement two of the major methods for generating new features: generating chunked representations by analyzing sets of objects into diagnostic components, and imprinting on frequently repeated complex stimuli. Both mechanisms seem to be necessary, and consequently, we do not advocate linking feature development to only one process, as **Burgund & Marsolek** apparently do when they associate adaptive feature learning with creating complex, holistic features but not with the componential processing of objects.

Some commentators (e.g., **Tijsseling**) argue that not only do we fail to provide a specific proposal for feature creation, but that such a proposal is impossible, in that features can be combined but not created (except through evolutionary time). There are strong existence proofs against this claim of impossibility. Pask's (1960) sound detection device (described in sect. R3.3 and by Cariani 1993) clearly increases its representational power by making pitch discriminations that it was originally incapable of making. Originally in this case, the device was structurally incapable of making pitch discriminations because it lacked the necessary physical links between electrodes. The process of physically extending filaments capable of conducting electricity altered the actual structure of the device, and hence its ability to register sound features.

The example shows that Fodor's (1980) argument (raised by **Braisby & Franks** and by **Tijsseling**) that it is impossible to develop a system with greater representational power than it had originally does not rule out the possibility of implementing feature creation. Physical alterations to a system ("traumas" in Fodor's framework) may change its representational capabilities and may be systematic and error-guided rather than accidental. Whereas Pask's discrimination learning device obviously changes its physical structure during training, the structural changes that accommodate human perceptual learning are perhaps more hidden, but certainly no less consequential. These structural changes include expansions in the regions of somatosensory cortex responsiveness to well-trained sound (Recanzone et al. 1993) and tactile (Recanzone et al. 1992) discriminations, and narrowing of receptive fields for neurons tuned to diagnostic aspects of a visual stimulus (Saarinen & Levi 1995). Like Pask's device, these neural mechanisms change the physical structure of the sensory system and can thereby generate representational changes that cannot be produced simply by recombining outputs obtained from existing primitives (Cariani 1993). Structural modification of sensory devices is not limited to an evolutionary time scale, as maintained by Tijsseling; it occurs instead within virtually every person's lifetime.

Finally, several commentators argue that positive proposals for feature learning are not possible because of an inherent circularity in the proposal (**Braisby & Franks, Huettel & Lockhead, Perruchet & Vinter**). **Braisby & Franks** suggest that our position is circular, because we use feature creation to explain categorization, but claim that categorization itself determines feature creation. We are indeed making the claim that feature learning is influenced by the categories we have as well as the more standard claim that category learning is influenced by the features we possess. Several neural networks implement exactly this kind of mutual influence. For example, in the S.O.S. network described by Goldstone et al. (in press), feature detectors adapt themselves to regions of a novel dimension and their adaptation is influenced by the categorization. Feature detectors become densely concentrated at category boundaries because the detectors in this region tend to miscategorize objects and thus send out "S.O.S." signals for other detectors to adapt themselves toward the trouble spots. Categorizations, in turn, depend solely on the outputs of the feature detectors. In other words, it is not necessary to first create detectors and then build associations between detectors and categories; both processes can operate in parallel, and should do so to model phenomena related to categorical perception (see also Harnad et al. 1991). In fact, many neural networks modify the nature of their internal units based on the backpropagation of errors from output units at the same time that the output units depend on the processing of the internal units (**Abdi et al., Dorffner, Hummel & Kellman**). Thus, the circularity perceived in our proposal can be thought of as bidirectional, and computable, interactions.

R6. Constraints on feature learning

In the target article, we maintained that not every stimulus aspect can be transformed into a psychological feature. Features are constrained by the task and by perception. In developing positive proposals for mechanisms that generate new features, several of the commentators elaborated on these constraints. These comments have led us to organize the constraints on feature creation in relation to: perception, unsupervised statistical information, supervised labels and categories, and task contexts.

R6.1. Perceptual constraints

Although we believe that newly created features are informed by task constraints, we certainly do not adopt an "anything goes" attitude, whereby any property can become a perceptual feature if it is required to accommodate a set of objects or tasks. The commentators present several proposals for specific perceptual constraints. **MacDorman** describes neurophysiological evidence that the visual system may organize the world into contiguous, oriented patches at different spatial resolutions – the outputs of Gabor filters. Given the nature of these filters, it would be difficult to create a highly coherent feature that encompassed widely separated and disjointed patches. Similarly, **Benson** provides evidence for a neurophysiological encoding of multidimensional convex hulls; such a coding scheme would again implement a strong bias for compact, contiguous features. **Singh & Landau** argue that procedures that find minima of curvature in shape outlines and create "cuts" to join these minima place constraints on the parts that are likely to be found.

The last of these constraints is particularly interesting because, as **Singh & Landau** note (see also sect. 1.2.2. of the target article), it can create part decompositions without assuming any initial analysis of a shape into specific detectors for these parts. Our only potential disagreement with these commentators is based on one interpretation of their claim that certain kinds of part-based categories may provide initial constraints on which learning can subsequently build. It is unlikely that perceptual constraints or innate categories first select an initial candidate set of parts that are then further pruned by learning. We believe that all these constraints act together, in parallel, to create the part features themselves. As argued in the target article, perceptual constraints by themselves underdetermine part segmentation, and constraints of the task alone underdetermine the process of reducing high-dimensional retinal inputs to low-dimensional part descriptions (see also Schyns & Murphy 1994).

At this stage, it is worth stressing that we advocate a continuum between interpretations and perceptual descriptions of objects. Interpretations of objects are highly influenced by knowledge, expectancies, and goals, whereas perceptual descriptions are not as labile, being constrained by our sensory transduction system. Two concentric circles may be interpreted as a bagel, but this does not necessarily imply that the perceptual features used to encode the circle are different from those used when they are interpreted as a doughnut (see **Tanaka**). Relative to interpretations, perceptual descriptions are more constrained by perceptual limitations, slower to change, more automatically processed, and less cognitively penetrable.

R6.2. Unsupervised statistical information

Some of the commentators were dissatisfied with our concentration on task constraints, arguing that even in the absence of strategic processing of stimuli according to a task or supervision, features can be acquired on the basis of the statistical properties inherent in the presented patterns themselves (**Cleeremans, Perruchet & Vinter, Phillips**). We did not intend to exclude the possibility that features could be acquired by picking up on regularities inherent in a set of stimuli, particularly if these regularities match the perceptual constraints of the system. We therefore welcome these comments as presenting evidence for additional bases of feature creation.

One reason to believe that unsupervised information provides an important source of constraints is that it occurs frequently. Stimuli rarely come with explicit labels attached, and so it makes good sense to look to the stimulus itself as a source of information to guide the imprinting of features. In fact, work with competitive learning algorithms has been important in showing that not only can units be formed that become specialized to particular inputs, but the major natural clusters inherent in a set of stimuli can often be determined without explicit feedback. Unsupervised learning is relatively underconstrained, however, and a large number of stimulus repetitions is required to acquire a relatively subtle feature if no external signal provides hints as to its existence, or if more salient features suggest alternative organizations for the set of stimuli. Via unsupervised learning, feature acquisition is typically a rather slow process, as is true of implicit learning devices (**Cleeremans, Perruchet & Vinter**).

Several recent proposals have attempted to increase the power of purely unsupervised learning, and in the process, have served to blur the line between supervised and unsupervised learning. De Sa and Ballard (1997) have shown that two sensory modalities that are trained at the same time and provide feedback for each other can reach a level of performance that would not be possible had they remained independent. [See also Ballard: "Diectic Codes for the Embodiment of Cognition" *BBS* 20(4) 1997.] This proposal can be extended to a single modality: acquisition of a subtle shape feature is possible if no explicit label is provided, if other shape features strongly suggest a category and thus supply implicit feedback. Hochberg (1997) has argued that feedback may not be as rare as it would appear. Humans tend to make eye fixations to the edge of objects to be recognized. While gaining experience with an object, every fixation that falls on a spatial location that is not part of the object can be taken as a source of feedback. Given the large number of fixations per minute, a substantial amount of feedback could guide a person to generate the correct pattern of fixations for tracing the edge of an object. In short, there are many proposals that increase the power of standard unsupervised learning without requiring the explicit labeling of standard supervised learning, thus establishing powerful techniques for perceptual adaptation.

R6.3. Supervised learning

Consistent with our claim that supervision is often necessary to guide feature development, several commentators elaborated on the role of supervised learning and labels on perceptual adaptation (**Abdi et al., Brooks, MacDorman, Quinn, Tanaka**). Brooks raises several points about the interactions between verbal labels and perceptual features: the same label may be used for different perceptual features, perceptual features are typically far richer than their labels, and labeling a feature can make it appear obvious when it would not otherwise have even been noticed. This constellation of results is compatible with the architecture described by **Grossberg.** During learning in the supervised ARTMAP system, the presence of the label leads to the formation of perceptual bundles that are diagnostic for the label. Presumably, presenting a label could activate a single unit, which could in turn activate the entire assembly of perceptual features to which the unit representing the label is connected.

Other commentators have helped refine the notion of supervised feedback. **Quinn** rightly points out that feedback involving a single category is much less effective than feedback on a set of contrasting categories as a guide to perceptual change. Echoing Gibson's (1969) claims, people tend to acquire distinctive features, and the distinctiveness ("diagnosticity" in our target article) of a feature within an object depends on the entire set of objects to be differentiated (**Lamberts**), even though there appear to be limitations on this principle depending on the order in which the objects are experienced (Schyns & Murphy 1994; Schyns & Rodet 1997). Quinn also challenges us to describe how supervised feedback interacts with perceptual constraints and object recognition processes to create new features. We have argued that features created by supervised feedback can be entered into the system's working vocabulary, and are available to be used by subsequent object recognition processes (Pevtzow & Goldstone 1994; Schyns & Murphy 1994). There are interesting experiments needed to decide whether particular types of task feedback are more or less important than particular perceptual constraints such as the Gestalt laws of organization (**Quinn**). However, it is generally not fruitful to consider these sources of constraints to be in opposition to each other. They are not mutually incompatible, and the supervised feedback that people receive in the real world is typically based on categories informed by perceptual constraints.

R6.4. Task contexts

In the target article we stressed the role of categorical relevance in developing features. This is one case in which the task confronting an organism leads to perceptual change, but the commentators have extended and differentiated this general notion of task dependency (**Deák, Dorffner, Lamberts, Tanaka**). The task confronting an organism is partially a function of the organism's history of categorization (or expertise, see Tanaka). The task during category learning may be different from the task required later, and Lamberts has shown that featural relevance in a categorization task may alter the perception of features, even after category learning has been completed. Deák, Lamberts, and **Hahn & Chater** all argue that perceptual biases change dynamically and rapidly, rather than being fixed aspects of the perceptual system.

It is also worth pointing out that constraints of learning, not simply perception, may also affect how features are formed. Although some commentators argue for very rapid changes to feature sets, as a function of the task (**Deák**), the set of alternative categories (**Lamberts**), and even the presentation of a single item (**Hahn & Chater**), **French & Weaver**'s skepticism with regard to fast perceptual changes is most likely justified. In almost all cases, for a feature to be added to the permanent vocabulary of a system, a substantial amount of training is required (see also **Cleeremans** and **Perruchet & Vinter**). Generally speaking, the longevity of a feature within a vocabulary will be proportional to its degree of exposure during training.

R6.5. Challenges for models of feature creation

The above survey of mechanisms and constraints should go a long way toward reducing the concern that feature creation is an unimplementable pipe dream. The problem has certainly not been solved, but the combination of proposals from robotics, computer science, and neurophysiology indicates that progress has been made in understanding some of the principles underlying perceptual adaptation and its uses. At the same time, we cannot point to a current system that we would hold up as an uncontroversial example of a device that genuinely creates novel, psychologically realistic features. We offer the following challenges, both to show how our proposal differs from existing computational models, and to inspire research into difficult but central issues concerning perception and learning.

R6.5.1. Perceptual constraints must be accommodated. The internal representations that are formed to guide categorization are shaped by the perceptual constraints mentioned in the previous section. Most network algorithms for creating features do not have biases that match psychophysical constraints (see **Grossberg**'s use of adaptive filters for an exception). As such, although backpropagation networks do create internal representations (**Dorffner, Hummel & Kellman**), there is more to feature creation than gradient descent on error. Furthermore, we do not believe that the strategy of having a fixed preprocessing of the input fed to an error-correction algorithm will go very far toward implementing the feature creation principles we discuss. Instead, perceptual constraints will need to be an intrinsic component of the computation of the error itself.

R6.5.2. Structured objects must be accommodated. Many effective models of feature learning create stimulus-wide features, sometimes called "holons" to reflect the fact that they are not restricted to a component of the object to be represented, but are instead spread across the entire object. PCA (**Abdi et al.**) and template-construction (**Edelman**) are examples of these techniques. It is a challenge for them to develop efficient representations for structured and articulated objects – objects that are clearly composed of individual parts standing in certain relation to each other. For example, a human body can adopt a wide variety of postures, from sitting to standing to running (and all postures in between). To recognize all of these postures as examples of human bodies it would be necessary to store a tremendous number of templates, and it is not clear whether any overall component of a PCA would be able to distinguish the various postures of a human body from those of a gorilla. An alternative approach is to represent bodies in terms of parts (e.g., arms, legs, chest, and head) and their relations to other parts. If a structural approach such as this were to incorporate learning as well, not only could holons be created, but so could local features of the object. With these local features, articulated objects with different relations between their parts could be correctly judged to be related. In sum, articulated objects illustrate an abstract challenge discussed in section R2: Template-based approaches need principles to go from holistic to componential representations, or to decompose their templates into their components.

R6.5.3. Evidential support from psychology. Currently, the most concrete proposals for feature creation mechanisms have come from computer science. Considerable work is still needed to test whether the features devised by these models are supported by evidence from cognitive psychology. To take PCA (**Abdi et al.**) as an example, it will be important to know whether people actually treat the statistically derived components as individuated components. Using the characterization of a psychological feature as an isolable unit of stimulus information, the components of PCA can be tested by observing (1) whether subjects can attend selectively to one component despite variations on the other components, and (2) whether they tend to separate components perceptually and then automatically recombine them in new arrangements. To the extent that the answer to these empirical questions is "yes," we would be more likely to believe that the components pulled out by the statistical PCA technique are psychologically valid.

R6.5.4. When are new features created? To the extent that they are viewed as creating feature detectors at their hidden units, standard backpropagation networks are constantly revising their feature sets slightly with every trial. Consistent with these networks, **Williams et al.** argue that every stimulus exposure does in fact alter featural descriptions. Alternatively, in **Grossberg**'s ARTMAP system, feature units are generated only if the mismatch between an incoming signal and the closest existing feature is sufficiently large. Deciding when a new feature is required for a stimulus set and when existing features provide adequate coverage is a complex process, depending on memory limitations, the similarity between existing and proposed features, and the requirements for representational precision. Unless features are incrementally adapted on every exposure to a stimulus, processes must not completely ignore small discrepancies between ideal and current features by regularizing the stimuli to fit the current feature set. Such an assimilation process could preserve a suboptimal feature set despite the occurrence of systematic violations of it. One possible solution to this problem is not to modify features on a moment-by-moment basis (a solution that could give rise to catastrophic forgetting of previously stored objects), but rather to increase the allowable range of variation along existing features. But then, when is a variation sufficiently large for a new feature to be created?

R7. Conclusions

"The main point to realize is that all knowledge presents itself within a conceptual framework adapted to account for previous experience and that any such frame may prove too narrow to comprehend new experiences." (Niels Bohr 1958).

The primary objective of our target article was to describe what we believe to be a missing link in categorization theories: the origin of features. We suggested that features could sometimes be created in response to new categorization demands. Future theories of category learning will need to include principles for feature creation similar to those discussed here. Many commentators offered constructive comments and specific proposals that helped to lay out the basic challenges that lie ahead. The fixed feature stance, however, is by no means the strawman pictured by **French & Weaver.** Several commentators (e.g., **Palmeri** and **Tijsseling**) argued that fixed features should form the basis of category learning theories until specific proposals for feature creation arise. On the contrary, we feel that much creative effort will be necessary to lay out the basic computational principles that will endow artificial systems with the sort of flexibility that appears to be necessary in real-life categorizations. Rather than shying away from the difficulty of the task, we should embrace its challenges.

References

Letters "a" and "r" appearing before authors' initials refer to target article and response respectively.

Abdi, H. (1988) Generalized approaches for connectionist auto-associative memories: Interpretation, implication, and illustration for face processing. In: *Artificial intelligence and cognitive science,* ed. J. Demongeot. Manchester University Press. [HA]

Abdi, H., Valentin, D., Edelman, B. & O'Toole, A. (1995) More about the difference between men and women: Evidence from linear neural networks and the principal component approach. *Perception* 24:539–62. [HA]

(1996) A Widrow-Hoff learning-rule for a generalization of the linear auto-association. *Journal of Mathematical Psychology* 40:175–82. [HA]

Abdi, H., Valentin, D. & O'Toole, A. (1997) A generalized auto-associator model for face semantic processing. In: *Optimality in biological and artificial networks?,* ed. D. Levine & W. R. Elsberry. Lawrence Erlbaum. [HA]

Aha, D. (1992) Tolerating noisy, irrelevant, and novel attributes in instance-based learning algorithms. *International Journal of Man-Machine Studies* 36:267–87. [UH]

Aha, D. W. & Goldstone, R. L. (1992) Concept learning and flexible weighting. *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society,* 534–39. Lawrence Erlbaum. [aPGS]

Ahissar, M. & Hochstein, S. (1977) Task difficulty and the specificity of perceptual learning. *Nature* 387:401–06. [PJB]

Algom, D. (1992) *Psychophysical approaches to cognition.* North Holland. [aPGS]

Anderson, J. A. & Mozer, M. (1981) Categorization and selective neurons. In: *Parallel models of associative memory,* ed. G. Hinton & J. Anderson. Erlbaum. [HA]

Anderson, J. A. & Rosenfeld, E. (1988) *Neurocomputing, foundations of research.* MIT Press. [aPGS]

Anderson, J. R. (1990) *The adaptive character of thought.* Erlbaum. [KW-H]

Andrews, J., Livingston, K., Harnad, S. & Fisher, U. (1997) Learned categorical perception in human subjects: Implications for symbol grounding (submitted). [aPGS]

Asfour, Y. R., Carpenter, G. A., Grossberg, S. & Lesher, G. (1993) Fusion ARTMAP: A neural network architecture for multi-channel data fusion and classification. In: *Proceedings of the World Congress on Neural Networks, vol. 2.* Erlbaum. [SG]

Ashby, F. G. (1992) Multidimensional models of categorization. In: *Multidimensional models of perceptions and cognition,* ed. F. G. Ashby. Erlbaum. [TJP]

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

(1980) The absolute efficiency of perceptual decisions. *Proceedings of the Royal Society of London [Biology]* 290:71–82. [aPGS]

Barlow, H. B. & Reeves, B. C. (1979) The versatility and absolute efficiency of detecting mirror symmetry in random dot displays. *Vision Research* 19:783–93. [aPGS]

Bates, E. & Elman, J. (1996) Learning rediscovered. *Science* 274:1849. [PP]

Bellman, R. E. (1961) *Adaptive control processes.* Princeton University Press. [aPGS]

Benson, P. J. (1995a) Perspectives on face perception: Directing research by exploiting emergent prototypes. In: *Cognitive and computational aspects of face recognition,* ed. T. Valentine. Routledge. [PJB]

(1995b) Unmasking the face: An algebraic approach. Paper presented at the Third Conference of the Australasian Cognitive Science Society. [PJB]

Berry, D. C. & Dienes, Z. (1993) *Implicit learning: Theoretical and empirical issues.* Erlbaum. [AC]

Biederman, I. (1987) Recognition-by-components: A theory of human image understanding. *Psychological Review* 94:115–47. [GD, CG, SG, JEH, PCQ, MS, aPGS, AT, JT]

Biederman, I. & Schiffrar, M. M. (1987) Sexing day-old chicks: A case study and expert systems analysis of a difficult perceptual-learning task. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:640–45. [JEH]

Bishop, C. M. (1995) *Neural networks for pattern recognition.* Clarendon Press. [GD]

Blakemore, C. & Over, R. (1974) Curvature detectors in human vision. *Perception* 3:3–7. [CRL]

Bohr, N. (1958) *Atomic physics and human knowledge.* Wiley. [rPGS]

Bongard, M. (1970) *Pattern recognition.* Spartan Books. [SAH]

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

Braine, M. D. S. (1994) Is nativism sufficient? *Journal of Child Language* 21:9–32. [GDe]

Braisby, N. R. & Franks, B. (1997) Why concepts only appear to be fuzzy (submitted). [NB]

Braisby, N. R., Franks, B. & Harris, J. (1996) On fuzzy perspectival approaches to classification. In: *Papers from CaRESSeS,* ed. D. Dubois. Editions Kime. [NB]

Bransford, J. D., Franks, J. J., Vye, N. J. & Sherwood, R. D. (1989) New approaches to instruction: Because wisdom can't be told. In: *Similarity and analogical reasoning,* ed. S. Vonsaidou & A. Orton. Cambridge University Press. [GDe]

Braunstein, M. L., Hoffman, D. D. & Saidpour, A. (1989) Parts of visual objects: An experimental test of the minima rule. *Perception* 18:817–26. [MS, aPGS]

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 & A. W. Young. Martinus Nijhoff. [aPGS]

Bruner, J. S., Goodnow, J. J. & Austin, G. A. (1956) *A study of thinking.* Wiley & Sons. [aPGS, AT]

Burns, B. & Shepp, B. E. (1988) Dimensional interactions and the structure of psychological space: The representation of hue, saturation, and brightness. *Perception and Psychophysics* 43:494–507. [aPGS]

Burt, P. & Adelson, E. H. (1983) The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications* 31:532–40. [aPGS]

Carey, S. (1985) *Conceptual change in childhood.* MIT Press. [aPGS]

(1991) Knowledge acquisition: Enrichment or conceptual change? In: *The epigenesis of mind,* ed. S. Carey & R. Gelman. Erlbaum. [aPGS]

(1992) Becoming a face expert. *Philosophical Transactions of the Royal Society of London, B* 335:95–103. [JT]

Cariani, P. (1993) To evolve an ear. *Systems Research* 10:19–33. [rPGS]

Carpenter, G. A. & Grossberg, S. (1991) *Pattern recognition by self-organizing neural networks.* MIT Press. [SG]

(1993) Normal and amnesic learning, recognition, and memory by a neural model of cortico-hippocampal interactions. *Trends in Neurosciences* 16:131–37. [SG]

Chalmers, D. J., French, R. M. & Hofstadter, D. R. (1992) High-level perception, representation and analogy. *Journal of Experimental and Theoretical Artificial Intelligence* 4:185–211. [aPGS]

Chapman, K. L., Leonard, L. B. & Mervis, C. B. (1986) The effect of feedback on young children's inappropriate word usage. *Journal of Child Language* 13:101–07. [aPGS]

Chater, N. & Hahn, U. (1997) Representational distortion and the universal law generalization (unpublished manuscript). [UH]

Clark, A. & Karmiloff-Smith, A. (1993) The cognizer's innards: A psychological and philosophical perspective on the development of thought. *Mind and Language* 8:487–519. [AC]

Clark, E. V. (1973) What's in a word? On the child's acquisition of semantics in his first language. In: *Cognitive development and the acquisition of language,* ed. T. E. Moore. Academic Press. [aPGS]

Cleeremans, A. (1993) *Mechanisms of implicit learning: Connectionist models of sequence processing.* MIT Press. [AC]

(1994a) Awareness and abstraction are graded dimensions. *Behavioral and Brain Sciences* 17:402–03. [AC]

(1994b) The representation of structure in explicit prediction tasks. In: *Attention and performance XV: Conscious and non-conscious information processing,* ed. C. Umiltà & M. Moscovitch. MIT Press. [AC]

(1997) Principles for implicit learning. In: *What is implicit about implicit learning,* ed. D. Berry. Oxford University Press. [AC]

Coren, S. & Ward, L. M. (1989) *Sensation and perception.* Harcourt Brace Jovanovich. [AT]

Crick, F. & Koch, C. (1995) Are we aware of neural activity in primary visual cortex? *Nature* 375:121–23. [PJB]

Crassini, B. & Over, R. (1975) Curvature-specific color after-effects. *Perception & Psychophysics* 17:398–404. [CRL]

Czerwinski, M., Lightfoot, N. & Shiffrin, R. M. (1992) Automatization and training in visual search. *American Journal of Psychology* 105:271–315. [aPGS]

Daugman, J. G. (1980) Two-dimensional spectral analysis of cortical receptive field profiles. *Vision Research* 20:847–56. [KFM]

(1985) Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America* 2(7):1160–69. [KFM]

Deák, G. (in preparation) Designing a paradigm for the study of semantic development: Replacing constraints with a scientifically viable framework. [GDe]

Deák, G. & Bauer, P. J. (1996) The dynamics of preschoolers' categorization choices. *Child Development* 67:740–67. [GDe]

Deák, G. O. (1995) The functional nature of induction: Flexibility in children's inferences about object properties. Unpublished doctoral thesis, University of Minnesota. [GDe]

De Groot, A. D. (1965) *Thought and choice in chess.* Mouton. [aPGS]

De Sa, V. R. & Ballard, D. H. (1997) Perceptual learning from cross-model feedback. In: *Pscyhology of learning and motivation, vol. 36,* ed. R. L. Goldstone, P. G. Schyns & D. L. Medin. Academic Press. [rPGS]

De Valois, R. L. & De Valois, K. K. (1990) *Spatial vision.* Oxford University Press. [aPGS]

Delk, J. L. & Fillenbaum, S. (1965) Differences in perceived color as a function of characteristic color. *American Journal of Psychology* 78:290–93. [aPGS]

Demartines, P. & Heraut, J. (1997) Curvilinear component analysis: A self-organizing neural network for nonlinear mapping of data sets. *IEEE Transactions on Neural Networks* 8:148–54. [aPGS]

Dickinson, D. K. (1989) Development of preschool children's ability to identify common materials. *Merrill-Palmer Quarterly* 35:165–80. [GDe]

Dominey, P. F. (1997) Reducing problem complexity by analogical transfer. *Behavioral and Brain Sciences* 20(1):71–72. [PFD]

Dominey, P. F, Ventre-Dominey, J., Brousolle, E. & Jeannerod, M. (1995) Analogical transfer in sequence learning: Human and neural-network models of fronto-striatal function. *Annals of the New York Academy of Sciences* 769:369–73. [PFD]

(1997) Analogical transfer is effective in a serial reaction time task in Parkinson's disease: Evidence for a dissociable sequence learning mechanism. *Neuropsychologia* 35:1–9. [PFD]

Dorffner, G. (1997) Radical connectionism – A neural bottom-up approach to AI. In: *Neural networks and a new AI,* ed. G. Dorffner. International Thompson Computer Press. [GD]

Dorffner, G., Hentze, M. & Thurner, G. (1996) A connectionist model of categorization and grounded word learning. In: *Proceedings of the Groningen Assembly on Language Acquisition (GALA '95),* ed. C. Koster & F. Wijnen. University of Groningen Press. [GD]

Driver, J. & Baylis, G. C. (1995) One-sided edge assignment in vision. 2. Part decomposition, shape description, and attention to objects. *Current Directions in Psychological Science* 4:201–06. [MS]

Edelman, S. (1995) Representation, similarity, and the chorus of prototypes. *Minds and machine* 5:45–68. [SE]

Edelman, S. & Bülthoff, H H. (1992) Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research* 32:2385–400. [aPGS]

Edelman, S. & Duvdevani-Bar, S. (1997) A model of visual recognition and categorization. *Philosophical Transactions of the Royal Society of London, Series B.* (in press). [SE]

Eimas, P. (1994) Categorization in early infancy and the continuity of development. *Cognition* 50:83–93. [aPGS]

Eimas, P. D. & Quinn, P. C. (1994) Studies on the formation of perceptually based basic-level categories in young infants. *Child Development* 65:903–17. [PCQ]

Eimas, P. D., Quinn, P. C. & Cowan, P. (1994) Development of exclusivity in perceptually based categories of young infants. *Journal of Experimental Child Psychology* 58:418–31. [PCQ]

Elio, R. & Anderson, J. R. (1981) The effects of category generalizations and instance similarity on schema abstraction. *Journal of Experimental Psychology: Human Learning and Memory* 7:397–417. [aPGS]

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D. & Plunkett, K. (1996) *Rethinking innateness.* MIT Press. [GD]

Estes, W. K. (1994) *Classification and cognition.* Oxford University Press. [TJP]

Fahle, M. & Morgan, M. (1996) No transfer of perceptual learning between similar stimuli in the same retinal position. *Current Biology* 6:292–97. [rPGS]

Fisher, D. L. (1986) Hierarchical models of visual search: Serial and parallel processing. Paper presented at the meeting of the Society for Mathematical Psychology, Cambridge, MA. [aPGS]

Foard, C. F. & Kemler Nelson, D. G. (1984) Holistic and analytic modes of processing: The multiple determinants of perceptual analysis. *Journal of Experimental Psychology: General* 113:94–111. [aPGS]

Fodor, J. A. (1975) *The language of thought.* Harvard University Press. [GDe]

(1980) On the impossibility of acquiring more powerful structures. In: *Language and learning,* ed. M. Piatelli-Palmarini. Routledge. [NB]

(1987) Modules, frames, pidgeons, sleeping dogs, and music of the spheres. In: *The robot's dilemma: The frame problem in artificial intelligence,* ed. Z. W. Pylyshyn. Ablex. [KFM]

(1994) Jerry A. Fodor. In: *A companion to the philosophy of mind,* ed. S. Guttenplan. Blackwell. [KFM]

Fodor, J. A. & Pylyshyn, Z. W. (1988) Connectionism and cognitive architecture: A critical analysis. *Cognition* 28:3–71. [aPGS]

Francis, G. & Grossberg, S. (1996) Cortical dynamics of boundary segmentation and reset: Persistence, afterimages, and residual traces. *Perception* 25:543–67. [SG]

Franks, B. (1995) Sense generation: A "quasi-classical" approach to concepts and concept combination. *Cognitive Science* 19:441–505. [NB]

Franks, B. & Braisby, N. R. (1997) On the relation between word use and concepts: Perspectives (submitted). [NB]

Franks, J. J. & Bransford, J. D. (1971) Abstraction of visual patterns. *Journal of Experimental Psychology* 90:65–74. [UH]

Frege, G. (1892/1966) On sense and reference. In: *Translations from the philosophical writings of Gottlieb Frege,* ed. P. Geach & M. Black. Blackwell. [NB]

French, R. M. & Hofstadter, D. (1991) Tabletop: An emergent, stochastic model of analogy-making. In: *Proceedings of the Thirteenth Annual Cognitive Science Society Conference,* 708–13. Erlbaum. [aPGS]

Friedman, J. H. & Tukey, J. W. (1974) A projection pursuit algorithm for exploratory data analysis. *IEEE Transactions on Computers* C, L3:881–89. [aPGS]

Fujita, L., Tanaka, K., Ito, M. & Cheng, K. (1992) Columns for visual features of objects in monkey inferotemporal cortex. *Nature* 360:343–46. [AT]
Garner, W. R. (1962) *Uncertainty and structure as psychological concepts.* Wiley. [SAH]
(1974) *The processing of information and structure.* Erlbaum. [SAH, aPGS, AT]
Gauker, C. (1994) *Thinking out loud: An essay on the relation between thought and language.* Princeton University Press. [CG]
Gauthier, I. & Tarr, M. J. (1997) Becoming a "Greeble" expert: Exploring the face recognition mechanism. *Vision Research* 37:1673–82. [PW]
Gauthier, I., Tarr, M. J., Anderson, P., Skudlarsky, P. & Gore, J. (1997a) Expertise training with novel objects can recruit the fusiform face area. *27th Annual Meeting of the Society for Neuroscience* (submitted). [PW]
Gauthier, I., Williams, P., Tarr, M. J. & Tanaka, J. W. (1997b) Training "Greeble" experts: A framework for studying expert object recognition processes (submitted). [PW]
Gelman, S. A. (1988) The development of induction within natural kind and artifact categories. *Cognition* 20:65–95. [aPGS]
Geman, S., Bienestock, E. & Doursat, R. (1992) Neural networks and the bias/variance dilemma. *Neural Computation* 4:1–58. [aPGS]
Gibson, E. J. (1969) *Principles of perceptual learning and development.* Appleton-Century Crofts. [JEH, CRL, arPGS]
(1971) Perceptual learning and the theory of word perception. *Cognitive Psychology* 2:351–68. [aPGS]
(1991) *An odyssey in learning and perception.* MIT Press. [aPGS, AT]
Gibson, E. J. & Walk, R. D. (1956) The effect of prolonged exposure to visually presented patterns on learning to discriminate them. *Journal of Comparative and Physiological Psychology* 49:239–42. [aPGS]
Gibson, J. J. & Gibson, E. J. (1955) Perceptual learning: Differentiation or enrichment? *Psychological Review* 62:32–41. [GDe]
Gluck, M. A. & Bower, G. H. (1988) Evaluating an adaptive network model of human learning. *Journal of Memory and Language* 27:166–95. [aPGS]
Goldstone, R. L. (1994a) Influences of categorization on perceptual discrimination. *Journal of Experimental Psychology: General* 123:178–200. [aPGS]
(1994b) The role of similarity in categorization: Providing a groundwork. *Cognition* 52:125–57. [aPGS]
(1995) Effects of categorization on color perception. *Psychological Science* 6:298–304. [aPGS]
Goldstone, R. L., Steyvers, M., Spencer-Smith, J. & Kersten, A. (in press) Interactions between perceptual and conceptual learning. In: *Cognitive dynamics: Conceptual change in humans and machines,* ed. E. Diettrich & A. Markman. MIT Press. [rPGS]
Goodman, N. (1965) *Fact, fiction, and forecast,* 2nd edition. Bobbs-Merrill. [aPGS]
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. [AT]
Grossberg, S. (1976) Adaptive pattern classification and universal recoding. *Biological Cybernetics* 21:145–59. [GD]
(1987) Cortical dynamics of three-dimensional form, color, and brightness perception: I. Monocular theory. *Perception and Psychophysics* 41:117–58. [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]
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. (1994) A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object recognition. *Psychological Review* 101:470–89. [SG]
(1997) Visual brain and visual perception: How does the cortex do perceptual grouping? *Trends in Neurosciences* 20:106–11. [SG]
Grossberg, S. & Stone, G. (1986) Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review* 93:46–74. [GD]
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. [SG]
Hahn, U. (1996) Cases and rules in categorization. Doctoral thesis, University of Oxford. [UH]
Hahn, U. & Chater, N. (1997) Concepts and similarities. In: *Knowledge, concepts and categories,* ed. L. Lamberts & D. Shanks. Psychology Press/MIT Press. [UH]
(in press) Understanding similarity: A joint project for psychology, case-based reasoning and law. *Artificial Intelligence Review.* [UH]
Hahn, U. & Vogel, C. (1997) Notes on logicism: Formalization and legal knowledge. *European Journal of Law, Philosophy and Computer Science* 1. [UH]
Hampton, J. A. (1987) Inheritance of attributes in natural concept conjunctions. *Memory and Cognition* 15:55–71. [NB]
Hancock, P. J., Burton, A. M. & Bruce, V. (1996) Face processing: Human perception and principal components analysis. *Memory and Cognition* 24:26–40. [HA]
Harnad, S. (1987) Category induction and representation. In: *Categorical perception: The groundwork of cognition,* ed. S. Harnad. Cambridge University Press. [GD, KFM, aPGS]
(1990) The symbol grounding problem. *Physica D* 42:335–46. [KFM, aPGS]
(1993) The frame problem as a symptom of the symbol grounding problem. *Psycoloquy* 4(34):frame-problem.11. [KFM]
Harnad, S., ed. (1987) *Categorical perception: The groundwork of cognition.* Cambridge University Press. [PJB]
Harnad, S., Hanson, S. J. & Lubin, J. (1991) Categorical perception and the evolution of supervised learning in neural nets. In: *Working papers of the AAAI spring symposium on machine learning of natural language and ontology,* ed. D. W. Powers & Reeker. [rPGS]
Helson, H. (1933) The fundamental propositions of Gestalt psychology. *Psychological Review:* 4:13–32. [PCQ]
Hill, H., Schyns, P. G. & Akamatsu, S. (1997) Information and viewpoint dependence in face recognition. *Cognition.* [aPGS]
Hinton, G., Williams, K. & Revow, M. (1992) Adaptive elastic models for handprinted character recognition. In: *Advances in neural information processing systems, IV,* ed. J. Moody, S. Hanson & R. Lippmann. [aPGS]
Hochberg, J. (1997) The affordances of perceptual inquiry: Pictures are learned from the world, and what that fact might mean about perception quite generally. In: *Psychology of learning and motivation, vol. 36,* ed. R. L. Goldstone, P. G. Schyns & D. L. Medin. Academic Press. [rPGS]
Hoffman, D. D. & Richards, W. A. (1984) Parts of recognition. *Cognition* 18:65–96. [MS, aPGS]
Hoffman, D. D. & Singh, M. (1997) Salience of visual parts. *Cognition* 63:29–78. [MS]
Homa, D. (1984) On the nature of categories. *Psychology of Learning and Motivation* 18:49–94. [TJP]
Hubel, D. H. & Wiesel, T. N. (1965) Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. *Journal of Neurophysiology* 28:229–89. [AT]
(1977) Ferrier lecture: Functional architecture of macaque monkey visual cortex. *Proceedings of the Royal Society of London, Biology* 198:1–59. [rPGS]
Humphrey, N. K. & Weiskrantz, L. (1971) Vision in monkeys after removal of the striate cortex. *Nature* 215:595–97. [PJB]
Hunt, E. & Agnoli, F. (1991) The Whorfian hypothesis: A cognitive psychology perspective. *Psychological Review* 98:377–89. [KW-H]
Imai, S. (1977) Pattern similarity and cognitive transformations. *Acta Psychologica* 41:433–47. [UH]
Intrator, N. (1992) Feature extraction using an unsupervised neural network. *Neural Computations* 4:98–107. [aPGS]
(1993) Combining exploratory projection pursuit and projection pursuit regression. *Neural Computation* 5:443–55. [aPGS]
Intrator, N. & Edelman, S. (1997) Learning low dimensional representations of visual objects with extensive use of prior knowledge. *Network* 8:259–81. [SE]
Intrator, N. & Gold, J. (1993) Three dimensional object recognition using an unsupervised BCM network: The usefulness of distinguishing features. *Neural Computation* 5:61–74. [aPGS]
Jacobson, R., Fant, G. & Halle, M. (1963) *Preliminaries to speech analysis: The distinctive features and their correlates.* MIT Press. [aPGS]
Jacoby, L. L., Baker, J. G. & Brooks, L. R. (1989) The priority of the specific episodic effects in picture identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 15:275–81. [LRB]
Janlert, L.-E. (1996) The frame problem: Freedom or stability? With pictures we can have both. In: *The robot's dilemma revisited,* ed. K. M. Ford & Z. W. Pylyshyn. Ablex. [KFM]
Jiménez, L., Méndez, C. & Cleeremans, A. (1996) Direct and indirect measures of implicit learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 22:948–69. [AC]
Jones, S. & Smith, L. (1993) The place of perception in children's concepts. *Cognitive Development* 8:113–39. [aPGS]
Julesz, B. (1981) Textons, the elements of texture perception, and their interactions. *Nature* 290:91–97. [JEH]
Kaas, J. H. (1995) The reorganization of sensory and motor maps in adult mammals. In: *The cognitive neurosciences,* ed. M. S. Gazzaniga. MIT Press. [aPGS]

Karni, A. & Sagi, D. (1991) Where practice makes perfect in texture discrimination: Evidence for primary visual cortex plasticity. *Proceedings of the National Academy of Sciences USA* 88:4966–70. [rPGS]

Katz, J. J. & Fodor, J. A. (1963) The structure of a semantic theory. *Language* 39:70–210. [aPGS]

Keil, F. C. (1989) *Concepts, kinds and cognitive development.* MIT Press. [aPGS]

Kennedy, J. (1993) *Drawing and the blind.* Yale University Press. [LRB]

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. [AT]

Kohonen, T. (1984) *Self-organization and associative memory.* Springer-Verlag. [aPGS]

Koza, J. R. (1994) *Genetic programming II.* MIT Press. [aPGS]

Kruschke, J. K. (1991) ALCOVE: A connectionist model of human category learning. In: *Advances in neural information processing 3,* ed. R. P. Lippmann et al. Morgan Kaufman. [GD]

(1992) ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review* 99:22–44. [TJP, aPGS]

Kurbat, M. A. (1994) Structural description theories: Is RBC/JIM a general purpose theory of human entry-level object recognition. *Perception* 23:1339–68. [aPGS]

Lamberts, K. & Chong, S. (1997) Rational models of categorization and flexible similarity. In: *Rational models of cognition,* ed. N. Chater & M. Oaksford. Oxford University Press. [KL]

Landau, B. (1994) Object shape, object name, and object kind: Representation and development. In: *Psychology of Learning and Motivaton, vol. 31,* ed. D. L. Medin. Academic Press. [MS, aPGS]

Landau, B., Smith, L. B. & Jones, S. S. (1988) The importance of shape in early lexical learning. *Cognitive Development* 3:299–321. [GDe, MS]

Latimer, C. R., Joung, W. & Stevens, C. J. (1994) Modelling symmetry detection with back-propagation networks. *Spatial Vision* 8:1–17. [CRL]

Lawrence, D. H. (1949) Acquired distinctiveness of cues: 1. Transfer between discriminations on the basis of familiarity with the stimulus. *Journal of Experimental Psychology* 39:770–84. [aPGS]

Li, M. & Vitani, P. (1993) *An introduction to Kolmogorov complexity theory and its applications.* Springer. [UH]

Livingstone, M. & Hubel, D. H. (1988) Segregation of form, color, movement and depth: Anatomy, physiology, and perception. *Science* 240:740–49. [AT]

MacDorman, K. F. (1997a) How to ground symbols adaptively. In: *Readings in computation, content and consciousness,* ed. S. O'Nuallain, P. McKevitt & E. MacAogain. John Benjamins. [KFM]

(1997b) *Symbol grounding: Learning categorical and sensorimotor predictions for coordination in autonomous robots.* Technical Report no. 423, Computer Laboratory, Cambridge University (e-mail librarian@cl.cam.ac.uk). [KFM]

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. [AT]

Malt, B. C. (1995) Category coherence in cross-cultural perspective. *Cognitive Psychology* 29:85–148. [KW-H]

Markman, E (1989) *Categorization and naming in children. Problems of induction.* MIT Press, Bradford Books. [aPGS]

(1995) Constraints on word meaning in early language acquisition. In: *The acquisition of the lexicon,* ed. L. Gleitman & B. Landau. MIT Press. [aPGS]

Marr, D. (1982) *Vision. A computational investigation into the human representation and processing of visual information.* W. H. Freeman. [MS, aPGS]

Marsolek, C. J. (1995) Abstract-visual-form representations in the left cerebral hemisphere. *Journal of Experimental Psychology: Human Perception and Performance* 21:375–86. [EDB]

Marsolek, C. J. & Burgund, E. D. (in press) Computational analyses and hemispheric asymmetries in visual-form processing. In: *Cerebral asymmetries in sensory and perceptual processing,* ed. S. Christman. Elsevier. [EDB]

Marsolek, C. J., Schacter, D. L. & Nicholas, C. D. (1996) Form-specific visual priming for new associations in the right cerebral hemisphere. *Memory and Cognition* 24:539–56. [EDB]

Matheus, C. J. (1991) The need for constructive induction. In: *Proceedings of the Eighth International Workshop on Machine Learning,* ed. L. Birnbaum & G. Collins. Morgan Kaufmann. [aPGS]

McGraw, G., Rehling, J. & Goldstone, R. L. (1994) Letter perception: Toward a conceptual approach. In: *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* 613–18. Erlbaum. [aPGS]

Medin, D. L. & Coley, J. D. (in press) Concepts and categorization. In: *Handbook of perception and cognition; Perception and cognition at century's end: History, philosophy, theory,* ed. J. Hochberg & J. E. Cutting. Academic Press. [MS]

Medin, D. L., Goldstone, R. L. & Gentner, D. (1993) Respects for similarity. *Psychological Review* 100:254–78. [aPGS]

Michalski, R. S. (1983) A theory and methodology of inductive learning. *Artificial Intelligence* 20:111–61. [aPGS]

Mitchell, M. (1993) *Analogy-making as perception.* MIT Press. [aPGS]

Moscovici, S. & Personnaz, B. (1991) Studies in social influence: VI. Is Lenin orange or red? Imagery and social influence. *European Journal of Social Psychology* 21:101–18. [aPGS]

Mozer, M. (1994) Computational approaches to functional feature learning. In: *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society,* 975–76. Erlbaum. [aPGS]

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

Murphy, G. L. & Ross, B. (1994) Predictions from uncertain categories. *Cognitive Psychology* 27:148–93. [aPGS]

Newell, A. & Rosenbloom, P. S. (1981) Mechanisms of skill acquisition and the law of practice. In: *Cognitive skills and their acquisition,* ed. J. R. Anderson. Erlbaum. [KW-H]

Newell, A. & Simon, H. A. (1972) *Human problem solving.* Prentice-Hall. [AC]

Norman, G. R., Le Blanc & Brooks, L. R. (1997) On the difficulty of noticing obvious features in patient appearance (submitted). [LRB]

Nosofsky, R. M. (1984) Choice, similarity and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10:104–14. [TJP]

(1987) Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 13:87–108. [aPGS]

(1992) Exemplar-based approach to relating categorization, identification, and recognition. In: *Multidimensional models of perceptions and cognition,* ed. F. G. Ashby. Erlbaum. [TJP]

Nosofsky, R. M., Palmeri, T. J. & McKinley, S. C. (1994) Rule-plus-exception model of classification learning. *Psychological Review* 101:53–79. [TJP, aPGS]

Oaksford, M. & Chater, N. (1991) Against logicist cognitive science. *Mind and Language* 6:1–78. [UH]

Oja, E. (1982) A simplified neuron model as a principal components analyzer. *Journal of Mathematical Biology* 15:267–73. [aPGS]

Oliva, A. & Schyns, P. G. (1995) Mandatory scale perception promotes flexible scene categorization. *Proceedings of the XVII Meeting of the Cognitive Science Society,* 159–63. Erlbaum. [aPGS]

(1997) Coarse tasks or fine edges? Evidence that information diagnosticity changes the perceptions of complex visual stimuli. *Cognitive Psychology* 34:72–107. [aPGS]

Olshausen, B. & Field, D. (1996) Natural image statistics and efficient coding. *Network* 7:333–39. [PJB]

O'Toole, A., Abdi, H., Deffenbacher, K. & Valentin, D. (1993) A low dimensional representation of faces in the hgher dimensions of the space. *Journal of the Optical Society of America A* 10:405–11. [HA]

O'Toole, A., Deffenbacher, K., Abdi, H. & Bartlett, J. (1991) Simulating the other race effect as a problem in perceptual learning. *Connection Sciences* 3:163–78. [HA]

O'Toole, A., Vetter, T., Troje, N. F. & Bülthoff, H. H. (1997) Sex classification is better with three-dimensional structure than with image intensity information. *Perception* 26:75–84. [HA]

Palmer, S. (1977) Hierarchical structure in perceptual representation. *Cognitive Psychology* 9:441–74. [aPGS, AT]

Palmeri, T. J. & Nosofsky, R M. (1997) Prototypes and ideal points in "ill-defined" perceptual categorization (submitted). [TJP]

Palmer, S., Rosch, E. & Chase, P. (1981) Canonical perspective and the perception of objects. In: *Attention and performances IX,* ed. J. Long & A. Baddeley. Erlbaum. [aPGS]

Partee, B. H. (1984) Compositionality. In: *Varieties of formal semantics,* ed. F. Landman & F. Veltman. Foris. [NB]

Pask, G. (1960) The natural history of networks. In: *Self-organizing systems,* ed. M. C. Yovits & S. Cameron. Pergamon Press. [rPGS]

Pasnak, R. (1971) Pattern complexity and response to distinctive features. *American Journal of Psychology* 84:235–45. [CRL]

Perruchet, P. (1997) A subjective unit formation account of implicit learning. In: *What is implicit about implicit learning,* ed. D. Berry. Oxford University Press. [AC, PP]

Perruchet, P. & Vinter, A. (1997) Learning and development: The implicit knowledge assumption reconsidered. In: *Handbook of implicit learning,* ed. M. Stadler & P. French. Sage Publications. [PP]

Perruchet, P., Vintner, A. & Gallego, J. (1997) Implicit learning shapes new conscious percepts and representations. *Psychonomic Bulletin and Review* 4:43–48. [PP]

Peterson, M. A. & Gibson, B. S. (1994) Must figure-ground organization precede

object recognition? An assumption in peril. *Psychological Science* 5:253–59. [aPGS]
Pevtzow, R. & Goldstone, R. L. (1994) Categorization and the parsing of objects. *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society.* Erlbaum. [PCQ, arPGS]
Phillips, W. A. (1974) On the distinction between sensory storage and short-term visual memory. *Perception and Psychophysics* 16:283–90. [WAP]
Phillips, W. A. & Singer, W. (1997) In search for common foundations for cortical computation. *Behavioral and Brain Sciences* 20(4):. [WAP]
Poggio, T. & Edelman, S. (1990) A network that learns to recognize three-dimensional objects. *Nature* 343:263–66. [aPGS]
Poggio, T., Fahle, M. & Edelman, S. (1992) Fast perceptual learning in visual hyperacuity. *Science* 256:1018–21. [SE]
Poggio, R. F. & von Gelder, T. (1995) *Mind as motion: Explorations in the dynamics of cognition.* MIT Press. [AC]
Pomerantz, J. R. & Lockhead, G. R. (1991) Perception of structure: An overview. In: *The perception of structure,* ed. G. R. Lockhead & J. R. Pomerantz. American Psychological Association. [SAH]
Posner, M. I. & Keele, S. W. (1968) On the genesis of abstract ideas. *Journal of Experimental Psychology* 77:353–63. [TJP]
Putnam, H. (1975) Philosophy and our mental life. In: *Mind, language and reality; Philosophical Papers, vol. 2.* Cambridge University Press. [JEH]
Quine, W. (1960) *Word and object.* MIT Press. [aPGS]
Quinn, P. C. (1987) The categorical representation of visual pattern information by young infants. *Cognition* 27:147–79. [PCQ]
Quinn, P. C., Brown, C. R. & Streppa, M. L. (1997) Perceptual organization of complex visual configurations by young infants. *Infant Behavior and Development* 20:35–46. [PCQ]
Quinn, P. C., Burke, S. & Rush, A. (1993) Part-whole perception in early infancy: Evidence for perceptual grouping produced by lightness similarity. *Infant Behavior and Development* 16:19–42. [PCQ]
Reber, A. S. (1993) *Implicit learning and tacit knowledge.* Oxford University Press. [AC]
Recanzone, G. H., Merzenich, M. M. & Jenkins, W. M. (1992) Frequency discrimination training engaging a restricted skin surface results in an emergence of a cutaneous response zone in cortical area 3a. *Journal of Neurophysiology* 67:1057–70. [rPGS]
Recanzone, G. H., Schreiner, C. E. & Merzenich, M. M. (1993) Plasticity in the frequency representation of primary auditory cortex following discrimination. *Journal of Neuroscience* 13:87–103. [rPGS]
Reiter, R. (1980) A logic for default reasoning. *Artificial Intelligence* 13:81–132. [UH]
Rescher, N. & Oppenheim, P. (1955) Logical analysis of gestalt concepts. *British Journal for the Philosophy of Science* 6:89–106. [CRL]
Riggs, L. A. (1973) Curvature as a feature of pattern vision. *Science* 181:1199–201. [CRL]
(1974) Curvature detectors in human vision? *Science* 184:1199–201. [CRL]
Rodet, L. & Schyns, P. G. (1994) Learning features of representation in conceptual context. In: *Proceedings of the XVI Meeting of the Cognitive Science Society,* 766–71. Erlbaum. [aPGS]
Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986) Learning internal representations by error propagation. In: *Parallel distributed processing: Explorations in the microstructure of cognition,* ed. J. L. McClelland & D. E. Rumelhart. Bradford Books. [GD, aPGS]
Saarinen, J. & Levi, D. M. (1995) Perceptual learning in Vernier acuity: What is learned? *Vision Research* 35:519–27. [rPGS]
Saffran, J. R., Newport, E. L., Aslin, R. N., Tunick, R. A. & Barrueco, S. (1997) Incidental language learning. *Psychological Science* 8:101–05. [PP]
Salzman, C. D., Britten, K. H. & Newsome, W. T. (1990) Cortical microstimulation influences perceptual judgements of motion direction. *Nature* 346:174–77. [AT]
Sammon, J. W. (1969) A nonlinear mapping algorithm for data structure analysis. *IEEE Transactions on Computers* C-18:401–09. [aPGS]
Sanger, T. D. (1989) Principal components, minor components, and linear neural networks. *Neural Networks* 2:459–73. [aPGS]
Schank, R. (1972) Conceptual dependency: A theory of natural language understanding. *Cognitive Psychology* 3:552–631. [aPGS]
Schyns, P. G. (1991) A neural network model of conceptual development. *Cognitive Science* 15:461–508. [aPGS]
(in press) Diagnostic recognition: Task constraints, object information and their interactions. *Cognition.* [rPGS]
Schyns, P. G. & Murphy, G. L. (1991) The ontogeny of units in object categories. In: *Proceedings of the XIII Meeting of the Cognitive Science Society,* 197–202. Erlbaum. [aPGS]
(1994) The ontogeny of part representation in object concepts. In: *The Psychology of Learning and Motivation,* ed. D. L. Medin. Academic Press. [MS, arPGS, JT]
Schyns, P. G. & Oliva, A. (1994) From blobs to boundary edges: Evidence for time and scale dependent scene recognition. *Psychological Science* 5:195–200. [aPGS]
Schyns, P. G. & Rodet, L. (1997) Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 23:681–96. [NB, PP, MS, arPGS, AT, KW-H]
Sejnowski, T. J., Kienkur, P. K. & Hinton, G. E. (1986) Learning symmetry groups with hidden units: Beyond the perceptron. *Physica* 22D:260–75. [CRL]
Sekuler, A. B., Palmer, S. E. & Flynn, C. (1992) Local and global processes in visual completion. *Psychological Science* 5:260–67. [aPGS]
Selfridge, O. G. (1959) Pandemonium: A paradigm for learning. In: *Symposium on the mechanization of thought process, vol. 1.* Proceedings of a Symposium held at the National Physical Laboratory, U.K., November 1958. Her Majesty's Stationery Office. [aPGS]
Serra, J. (1982) *Image analysis and mathematical morphology, vol. 1.* Academic Press. [aPGS]
Servan-Schreiber, D. & Anderson, J. R. (1990) Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16:592–608. [PP]
Seyranian, G., Singh, M. & Hoffman, D. D. (1997) Cuts for parsing visual shapes: 2. Experiments. *University of California, Irvine, Mathematical Behavioral Sciences Memo 97–03.* [MS]
Shashua, A. & Ullman, S. (1988) Structural saliency: The detection of globally salient structures using a locally connected network. *Proceedings of the International Conference on Computer Vision,* 321–27. [rPGS]
Shepard, R. N. (1957) Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika* 22:325–45 [aPGS]
(1989) Internal representations of universal regularities: A challenge for connectionism. In: *Neural connections, mental computation,* ed. L. Nadel, L. A. Cooper, P. Culicover & R. M. Harnish. MIT Press. [aPGS]
Singer, W. (1995) Developments and plasticity of cortical processing architectures. *Science* 270:758–64. [WAP]
Singh, M., Seyranian, G. & Hoffman, D. D. (1996) Cuts for parsing visual shapes: 1. Theory. *University of California, Irvine, Mathematical Behavioral Sciences Memo* 96–33. [MS]
Sirovich, L. & Kirby, M (1987) Low dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A* 4:519–24. [HA]
Smith, C., Carey, S. & Wiser, M. (1985) On differentiation: A case study of the development of the concepts of size, weight and density. *Cognition* 21:177–237. [aPGS]
Smith, L. B. & Kemler, D. G. (1978) Levels of experienced dimensionality in children and adults. *Cognitive Psychology* 10:502–32. [aPGS]
Snippe, H. P. & Koenderink, J. J. (1992) Discrimination thresholds for channel-coded systems. *Biological Cybernetics* 66:543–51. [AT]
Spelke, E. (1994) Initial knowledge: Six suggestions. *Cognition* 50:431–45. [aPGS]
Stoerig, P. & Cowey, A. (1997) Blindsight in man and monkey. *Brain* 120:535–59. [PJB]
Tanaka, J. W. & Sengco, J. A. (in press) Features and their configuration in face recognition. *Memory and Cognition.* [JT, PW]
Tanaka, J. W. & Taylor, M. (1991) Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology* 23:457–82. [arPGS, JT]
Tarr, M. J. & Pinker, S. (1989) Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology* 21:233–82. [aPGS]
Thibaut, J. P. (1994) Role of variation and knowledge on stimuli segmentation: Developmental aspects. Paper presented at the Sixteenth Annual Meeting of the Cognitive Science Society, Atlanta. [aPGS]
(1995) The development of features in children and adults: The case of visual stimuli. In: *Proceedings of the Seventeenth Annual Meeting of the Cognitive Science Society,* 194–99. Erlbaum. [aPGS]
Thibaut, J. P. & Schyns, P. G. (1995) The development of feature spaces for similarity and categorization. *Psychologica Belgica* 35:167–85. [aPGS]
Treisman, A. & Gelade, G. (1980) A feature-integration theory of attention. *Cognitive Psychology* 12:97–136. [aPGS]
Triesman, M., Faulkner, A., Naish, P. L. N. & Rosner, B. S. (1995) Voice-onset time and tone-onset time: The role of criterion-setting mechanisms in categorical perception. *Quarterly Journal of Experimental Psychology* 48A:334–66. [PJB]
Turk, M. & Pentland, A. (1991) Eigenfaces for recognition. *Journal of Cognitive Neurosciences* 3:71–86. [HA]
Tversky, B. & Hemenway, K. (1984) Objects, parts, and categories. *Journal of Experimental Psychology: General* 113:169–93. [JT]

Ullman, S. (1984) Visual routines. *Cognition* 18:97–159. [arPGS]
(1989) Aligning pictorial descriptions: An approach to object recognition. *Cognition* 32:193–254. [aPGS]
Valentin, D. (1996) "How come when you turn your head I still know who you are: Evidence from computational simulations and human behavior." Ph.D. dissertation, The University of Texas at Dallas. [HA]
Valentin, D. & Abdi, H. (1996) Can a linear auto-associator recognize faces from new orientations? *Journal of the Optical Society of America A* 13:717–24. [HA]
Valentin, D., Abdi, H. & Edelman, B. (1996) Connectionist "face-off": Different algorithms for different tasks. *Psychologica Belgica* 36:65–92. [HA]
Valentin, D., Abdi, H., Edelman, B. & O'Toole, A. J. (in press) Principal component and neural network analyses of face images: How stable is gender information? *Journal of Mathematical Psychology.* [HA]
Valiant, L. G. (1984) A theory of the learnable. *Communications of the ACM* 27:1134–42. [GD]
von der Marlsburg, C. (1973) Self-organization of orientation-sensitive cells in the striata cortex. *Kybernetik* 14:85–100. [rPGS]
Ward, T. B. (1983) Response tempo and separable-integral responding: Evidence for an integral-to-separable processing sequencing in visual perception. *Journal of Experimental Psychology: Human Perception and Performance* 9:103–12. [aPGS]
Watt, R. (1987) Scanning from coarse to fine spatial scales in the human visual system after the onset of a stimulus. *Journal of the Optical Society of America* A4:2006–21. [aPGS]
Weinberger, N. M. (1993) Learning-induced changes of auditory receptive fields. *Current Opinion in Neurobiology* 3:570–77. [rPGS]
Widrow, G. & Hoff, M. E. (1960) Adaptive switching circuits. *Institute of Radio Engineers, Western Electronic Show and Convention, Convention records, vol. 4,* 96–194. [aPGS]
Wisniewski, E. J. & Medin, D. L. (1994) On the interaction of theory and data in concept learning. *Cognitive Science* 18:221–81. [aPGS]
Witkin, A. (1986) Scale-space filtering. In: *Proceedings of the Ninth International Joint Conference on Artificial Intelligence,* 1019–22. Morgan Kaufmann. [aPGS]
Yulle, A. L. & Ullman, S. (1990) Computational theories of low-level vision. In: *Visual cognition and action, vol. 2,* ed. D. N. Osherson, S. M. Kosslyn & J. M. Hollerbach. MIT Press. [AT]