Diagnostic recognition: task constraints, object information, and their interactions

Philippe G. Schyns*

Department of Psychology 56, Hillhead St., University of Glasgow, Glasgow G12 8QB, UK

Abstract

Object recognition and categorization research are both concerned with understanding how input information matches object information in memory. It is therefore surprising that these two fields have evolved independently, without much cross-fertilization. It is the main objective of this paper to lay out the basis of a dialogue between object recognition and categorization research, with the hope of raising issues that could cross-fertilize both domains. To this end, the paper develops diagnostic recognition, a framework which formulates recognition performance as an interaction of task constraints and object information. I argue and present examples suggesting that diagnostic recognition could be fruitfully applied to the understanding of everyday object recognition. Issues are raised regarding the psychological status of the interactions specified in the framework. © 1998 Elsevier Science B.V. All rights reserved

Keywords: Object recognition; Categorization research; Diagnostic recognition

1. Introduction

Object recognition and categorization research are both concerned with the question ‘what is this object?’ To recognize an object as a car is not very different from placing the object in the car category. In both cases, the problem is to understand how input information matches with information in memory. Thus, both categorization and object recognition research are concerned with the same fundamental issues of ‘which input information should be used?’ ‘what is the organization of information in memory?’ and ‘how do input and memory information interact to explain performance?’

* E-mail: philippe@psy.gla.ac.uk

0010-0277/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved

PII S0010-0277(98)00016-X
Given such profound similarity, it is surprising that object recognition and categorization research have evolved separately, without much cross-fertilization between the two. The reason for this could be a difference of focus: typical categorization studies have sought to understand the rules governing the formation of categories (the idea that the visual attributes feathers, wings, legs, beak, black, but also functional attributes such as fly, lay eggs, live in the trees represent a crow which is also a bird, an animal and a living thing), while recognition researchers have mostly looked into the perceptual characteristics of the recognition process (e.g. the representation of the visual attributes of a crow that authorize its initial recognition as a member of the bird category). However, recent debates on the possible interactions between categorization and perception have suggested that the principles governing the formation of categories should be more tightly coupled with the perceptual aspects of recognition (Schyns et al., 1998). It is proposed that such interactions will promote the emergence of new, more powerful theories of visual cognition.

It is the main objective of this paper to lay out the basis of a dialogue between object recognition and categorization studies, to raise issues that could cross-fertilize both fields. To this end, the first section develops diagnostic recognition, a framework which integrates two main factors: the task constraints of categorization studies, and the perceptual information of recognition theories. I wish to stress from the outset that diagnostic recognition is not a new theory of object recognition. Instead, it is a generic framework which proposes one possible answer to the question: ‘how could we frame issues common to object recognition and categorization theories?’

The paper is organized as follows: the first section develops the diagnostic recognition framework and its main motivations. Two examples are then presented to illustrate how both object recognition and categorization could benefit from the framework. Subsequent sections apply diagnostic recognition to the explanation of ‘everyday recognition’, discuss its task constraints, its perceptual information, and the psychological status of their interactions.

2. The diagnostic recognition framework: interactions of task constraints and object information

Everyday observation reveals that a single object fits into many possible categories. For example, the same object may be recognized as a Porsche, a car, or a vehicle. On other occasions, it may be called a toy, an expensive gift, a public nuisance, or a public danger which sometimes leads to scrap metal. Categorization is a flexible process and people tend to place a single object into one category or another depending on a number of factors including their goals and actions, environmental contingencies, and so forth (e.g. Barsalou, 1983).

It is always worth stressing that different categorizations of an identical object tend to change the information requirements of the task at hand. For example, when assigning a visual event to the Porsche, collie, sparrow, Mary or New York category comparatively more specific information might be necessary than when categorizing...
it as a car, dog, bird, human face or city. I will not consider here all possible object classifications, but instead focus on the information requirements associated with the hierarchical organization of categories; the idea that an object belongs to a sequence of progressively more inclusive categories such as Porsche, car, vehicle. Within this hierarchy, I will concentrate on the initial, or so-called ‘perceptual’ classifications (e.g. Porsche or car), instead of the abstract functional classifications (e.g. vehicle) which are arguably more detached from the perceptual input. Henceforth, task constraints relate to the visual information required to place the input into the hierarchy of perceptual categories. Section 3.1 will detail what these task constraints might be. For now, it suffices to treat them as the demands for perceptual information that emanate from different categorizations. Although task constraints have traditionally been the province of categorization research, they are an irreducible factor of any recognition task, and the first factor considered in diagnostic recognition.

The second factor is the structure of the perceptual information available to form hierarchically organized categories. Objects form categories because they ‘look alike’ – i.e. they share cues such as a similar silhouette or global shape, distinctive sets of parts similarly organized (e.g. nose, mouth, eyes, ears, hair and their structural relationships), or characteristic surface properties (e.g. smooth vs. discontinuous, symmetric vs. asymmetric, and textural, color and illumination cues). Generally speaking, there are perceptual limitations to the extraction of image cues. For instance, the neural wiring of early vision could promote an earlier recovery of shape from luminance variations and shading, at the expense of color and textural cues (e.g. Livingstone and Hubel, 1987; Weisstein et al., 1992). Alternatively, real-world objects could be so structured that their shapes vary much more than their colors and textures, resulting in comparatively more information for visual processing along the shape dimension. Issues of perceptual availability of object information have traditionally been the province of perceptually-oriented object recognition researchers. However, perceptual cues are an irreducible factor of any object categorization, and the second factor of diagnostic recognition. Section 3.3 will review how the main theories of human object recognition assume that we use different kinds of perceptual information.

Diagnostic recognition attempts to frame a recognition problem as an interaction between task constraints and perceptual object information, the two factors just discussed (see Fig. 1). Here is how it would work: when the information required to assign an object to a category matches with input information, a subset of object cues become particularly useful (i.e. diagnostic) for the task at hand. Diagnosticity is the first component of recognition performance. However, perceptual limitations on the extraction of diagnostic cues should also affect performance. Thus, diagnostic recognition frames explanations of performance as interactions between cue diagnosticity and cue availability. The nature and the implications of these interactions have been neglected both in object recognition and in object categorization research.

For example, even though the diagnosticity of cues for different classifications has been thoroughly modeled in categorization theories (e.g. Elio and Anderson, 1981; Nosofsky, 1984, 1986; Estes, 1986; Gluck and Bower, 1988; Anderson, 1991;
Kruschke, 1992; and many others), these often adopt a stance of: "you tell me what the object cues are, and I will tell you how they are integrated to perform the object categorization" (Schyns et al., 1998). Consequently, they place few perceptual constraints on what may count as an object cue and they tend not to incorporate perceptual limitations on cue availability in their explanations of performance. However, recent studies have suggested that the simple fact of placing an object in one category or another could change the perception of its cues (Goldstone, 1994), or even its analysis into perceptual cues (Schyns and Murphy, 1991; Schyns and Rodet, 1997). If categorization does not start where perception ends, constraints on the perceptual availability of visual cues should supplement explanations of even very simple categorizations.

In contrast, whereas recognition researchers are well aware of the role of cue availability in their explanations of performance, they still tend to overlook the influence of task constraints. For example, researchers often assume that shape cues always supersede color and textural cues in everyday recognition (e.g. Biederman, 1987; Biederman and Ju, 1988; Tarr and Kriegman, 1998). Cutzu and Edelman (1997) even argued that the visual system could support ‘veridical’ representations of the objective similarities existing between the shapes of distal three-dimensional (3D) objects. These findings open the possibility of veridical, exhaustive representations of the object shapes themselves. However, would these always be exhaustive, irrespective of the diagnosticity of other cues? If this was the case, people who would distinguish different fruits in a basket on the basis of their colors and textures would also construct complete representations of their shapes. Alternatively, the diagnosticity of color and textural cues in this task could affect the accuracy of shape representations. People who categorize a chair as chair or Art Deco chair might also construct slightly different representations of its shape. Cutzu and Edelman’s findings suggest that the shape of 3D objects can sometimes be represented in
detail. The circumstances of exhaustive or task-dependent representations remain to be explored. These issues arise from considering the potential impact of cue diagnosticity on object perceptions and representations in recognition research.

Together, these considerations should justify the importance of bridging the gap between object categorization and object recognition research. A complete object recognition and categorization theory will need to integrate the factors affecting the diagnosticity of object cues and the perceptual constraints on their availability. It is the first aim of this paper to illustrate that the interactions between cue diagnosticity and cue availability could account for a wide range of phenomena (including feature learning, viewpoint-dependence vs. independence, basic and subordinate level recognition, the basic-to-subordinate shift, and scale-dependent recognition).

The second aim of the paper is to address the psychological status of the interactions between the task and the object cues. If several object cues are highly diagnostic of a categorization, their perception is essential to accomplish the task. What happens, then, to the other, less informative cues? Are they all perceived, or are some of them left out of the perception? When perceived, are non-diagnostic cues processed as extensively as when they are diagnostic? In the example discussed earlier, would one encode all the complex arabesques of a 3D Art Deco chair when one simply categorizes it as a chair? Simply put, is object perception driven to a certain extent (yet to be determined) by the diagnosticity of cues in a task, or does it reflect all the information of the distal object? There are psychological (O’Regan, 1992; Rensink et al., 1997; Schyns and Oliva, 1997) philosophical (Dennett, 1991) and computational (Cutzu and Edelman, 1997) indications that an exhaustive view is misguided. Later in the paper, I will present an instance of partial, flexible and diagnosticity-driven perceptions of complex visual stimuli. I will also argue that flexible perceptions could arise as psychological byproducts of the interactions between cue diagnosticity and cue availability. For the time being, let us turn to an important object recognition phenomenon (viewpoint-dependence) and examine how it might be fruitfully expressed within the diagnostic recognition framework.

2.1. Example 1: diagnostic recognition and viewpoint-dependence

One of the most challenging problems of object recognition is to explain the relative invariance of recognition to changes in object orientation. This is not to say that object recognition is fully viewpoint-invariant; there are now many independent sources of evidence suggesting that a large number of object classes are better recognized when shown from particular viewpoints (e.g. Palmer et al., 1981; Rock and Di Vita, 1987; Tarr and Pinker, 1989; Perrett et al., 1991; Bülthoff and Edelman, 1992; Edelman and Bülthoff, 1992; Vetter et al., 1994). Subjects label such object views as ‘better’ and are faster to categorize the objects shown in these views. Typically, ‘viewpoint-dependent recognition’ refers to a monotonic increase in recognition performance (reaction times and/or error rates) with increasing mis-orientation from the preferred views. Evidence of such viewpoint-dependent recognition has been reported for familiar (e.g. Palmer et al., 1981), unfamiliar (e.g. Rock and Di Vita, 1987), realistic and artificial objects (e.g. Tarr and Pinker,
1989; Bülthoff and Edelman, 1992), and the detailed conditions of viewpoint-depene-
dence have become a central issue in object recognition research. However, the debate is stil
open as to the interpretation of the phenomenon. Evidence of viewpoint-dependence or viewpoint-independen-
cess is often used to tease apart formats of object representation (see Biederman and Ger-

Fig. 2. This figure (adapted from Hill et al., 1997) illustrates the main results of an experiment on
viewpoint-dependence in face identification. The top left picture shows the face views of the shape-
from-shading condition. The bottom left histogram presents patterns of viewpoint-dependent performance
for the identification of shaded stimuli. The top right picture shows the face views of the shape plus texture
conditions, and the bottom right histogram presents the viewpoint-dependent performance for the identi-

1 The precise conditions for viewpoint invariance specified in Biederman and Gerhardstein (1993) are
listed in Section 3.3.
interaction between the diagnosticity of cues in a task and the availability of those cues in the input image. The presence of diagnostic information in these views, rather than the specific format (part or view) used to represent this information in memory, could then determine dependence of recognition on one (or a subset of) view(s). For example, two different categorizations of an identical face might change the information requirement of the task, the face cues that are diagnostic, and the subset of views that are preferred for recognition. To illustrate, many views would convey sufficient information to categorize Mary’s face as face. However, fewer views would allow one to classify the same face as Mary (because Mary’s features are visible from only a restricted subset of views). Within the range of views in which diagnostic cues are visible, there could be geometrical and perceptual limitations on their extraction from the image. For example, although the nose of a face is visible in all views between the two profiles, its length (if it was important to identify Mary) might be easier to measure from a 3/4, or profile view than from the frontal view. These limitations, interacting with task constraints, would predict that length-of-nose (in fact, its perceptual implementation) would only affect viewpoint-dependence when this cue was diagnostic for the task – i.e. in the Mary, not the face categorization.

Hill et al. (1997) tested the possibility of such relative patterns of viewpoint-dependent face recognition. Their subjects learned only one view of a face (either the full-face, FF, one three-quarter, TQ, or one profile view, PR, see Fig. 2) and were subsequently tested on their generalization capabilities to other views of the same face (FF, the two TQ and the two PR views). The top pictures of Fig. 2 illustrate what these views looked like. Different subject groups were assigned a different learning view, and all groups participated in the same generalization task which tested all views. This allowed a comparison of patterns of viewpoint-dependence when people extracted object information from a single, different view, but had to recognize all views. One set of experiments used shaded models of 3D laser-scanned faces, to isolate the influence of shape-from-shading cues (see Fig. 2, the top left picture). A second set of experiments added color and texture to the shaded models, to examine the role of these supplementary cues (see Fig. 2, the top right picture). Together, these designs framed viewpoint-dependence as an interaction between fixed task demands and variable stimulus information.

As an object class, faces share geometrical information to which perception could be attuned. One such property is their approximate bilateral symmetry (Vetter et al., 1994) which allows occluded cues to be inferred from a single learned view. Consequently, the learned view and its symmetrical view might be identified with equal accuracy, and possibly better than any other unseen views. Such effects of symmetric object information should be particularly salient with shaded face models, for which no other cue than shape is available from the image (compare the top pictures of Fig. 2).

The results of Hill et al. (1997) are summarized in the histograms of Fig. 2. The bottom left histogram illustrates that subjects who learned a TQ view recognized almost as efficiently the symmetric three-quarter, while performance decreased monotonically with rotation in depth for subjects who learned the profile, or the
full-face view, confirming the role of symmetric face information in explanations of viewpoint-dependent identification, or that the TQ view conveys more face information for identity decisions (see also Schyns and Bültzoff, 1994; Troje and Bültzoff, 1996).

The addition of color and textural cues in an otherwise identical task affected performance (see Fig. 2, bottom right histogram). It was found that learning a TQ view now elicited good generalization to all views (of those tested). Also, a symmetric peak appeared to the other profile when learning a colored profile. Color and textural cues offered supplementary object information which reduced the overall viewpoint-dependence for identity. Together, these two experiments illustrate how a different availability of object cues can change patterns of viewpoint-dependence.

Diagnostic recognition also predicts that the task and its associated cue diagnosticity could affect viewpoint-dependence. Pilot studies were run in a paradigm identical to Hill et al. (1997), with only one variation: the task. Rather than judging whether or not the learned and tested view were of the same person, subjects were now asked to judge whether they had the same gender. With shaded faces, performance was close to chance, but it was near ceiling with textured faces, with no marked dependence on viewpoint in either case. This contrasted with the viewpoint-dependent performance observed for identity judgments.

In sum, this example suggests that viewpoint-dependence can be framed as an interaction between the multiple categorizations of an object and its perceptual information. If a categorization requires selective input information, and if its extraction depends on viewpoint, recognition performance might reflect the requirement of ‘getting a good view’ of the diagnostic cues. Evidence of such preferred views could prompt an object recognition researcher to hypothesize that these views (or their information content) actually represent faces in memory. However, there would be a difficulty with this strategy if each change in task elicited a change in preferred views. We come back to this point in the General Discussion, in Section 5.2. The next example examines the impact of cue availability on explanations of even very simple categorizations.

### 2.2. Example 2: object information and categorization

As explained earlier, categorization research as a whole has underplayed the role of perceptual factors in its theories. In a typical categorization experiment, subjects are instructed to learn the rules to categorize simple colored geometric shapes (see e.g. Bruner et al., 1956; Shepard et al., 1961; Bourne, 1982). For example subjects could learn that ‘green and square’ defined the objects of a category. Importantly, there is no ambiguity as to which features characterize which objects. Although categorization research has considerably evolved since these pioneering experiments, tight control of stimulus dimensions is still required (e.g. Murphy and Smith, 1982; Gluck and Bower, 1988; Wattenmaker, 1991; among many others). Modern category learning models still adhere to a similar approach: They specify a number of dimensions along which the stimuli can vary, and these form the basis of the similarity comparisons which underlie category learning (see,

The idea that categorization processes operate on a ‘pre-perceived’ input has led researchers to concentrate comparatively more on the ways in which object cues combine to represent categories than on the origin of the cues themselves. However, it is legitimate to question whether object cues are fixed and independent of the categorization being performed, or whether they can flexibly tune to the perceptual characteristics of the object categories they must differentiate.

This issue comes into sharper focus once the perceptual availability of object cues is taken seriously. The medical expertise literature abounds with illustrations of category learning that do not seem to fit well with the idea that relevant cues are a given. For example, when complete novices categorize dermatosis (Norman et al., 1992), sex chicks (Biederman and Shiffrar, 1987) and read chest X-rays (Christensen et al., 1981; Lesgold, 1984) they are not always able to see the relevant cues of the stimuli. Expertise with these categories involves as much learning which cue goes with which categories as learning the object cues themselves.

Object cues are the first inputs to categorization; they provide a perceptual organization of the stimulus on which categorization processes operate. One function of object cues is therefore to represent important perceptual differences between categories. Reasoning backwards from this property, Schyns and Murphy (1994) suggested that the need to distinguish categories that initially ‘look alike’ could prompt the creation of new object cues that change the perception of the stimuli. The Functionality Principle summarizes this view (Schyns and Murphy, 1994): ‘If a fragment of a stimulus categorizes objects (distinguishes members from non-members), the fragment is instantiated as a unit in the representational code of object concepts’ (p. 310).

Schyns and Rodet (1997) tested one implication of the Functionality Principle, that orthogonal categorizations of the same stimulus could arise from its perceptual organization with different object cues. They reasoned that a different history of categorization of unfamiliar objects could change the cues people learn to perceptually organize the visual input. Their experiments involved categories of unfamiliar objects called ‘Martian cells’ (examples of cells are presented in Fig. 3). Not only were these objects unfamiliar to subjects, but their cues were also unfamiliar. Learning to categorize the cells involved as much learning which cues goes with which category as the cues themselves.

Categories were defined by specific blobs common to all members to which irrelevant blobs were added (to simulate various cell bodies). X cells shared the x cue, Y exemplars shared y, and the components x and y were always adjacent to one another in XY cells. (Fig. 3 shows, from left to right, an XY, an X, and a Y exemplar. It also shows their defining xy, x and y cues.). A difference in categorization history simply resulted from one group learning X before Y before XY (X → Y → XY) while the other group learned the same three categories, but in a different order (XY → X → Y). The idea was that this simple difference would elicit orthogonal perceptions and representations of the identical XY Martian cells.
Results revealed that $X \rightarrow Y \rightarrow XY$ subjects initially created the cues $x$ and $y$ when they learned their $X$ and $Y$ categories, respectively. The incoming $XY$ category was then perceived and represented as a conjunction of the acquired $x$ and $y$ cues. Cue creation was different in the group initially exposed to the $XY$ category. Unlike the other group, when $XY \rightarrow X \rightarrow Y$ subjects initially learned $XY$, they did not possess the $x$ and $y$ components that allowed a conjunctive analysis. Instead, subjects learned to perceive and represent $XY$ with a configural cue (that we call $xy$, but whose perceptual status is really more like an independent $z$ unit) without even noticing the $x \& y$ conjunction that the other group perceived.

This example illustrates that one cannot simply assume the cues on which classification processes operate. A simple change in the history of categorization of unfamiliar materials changed the cues that were learned, the perceptual analyses, perceptions and representations of identical objects. Because object cues form the basis of the similarity judgments that determine category learning, complete explanations of categorization behavior will need to integrate cue availability.

In summary of the examples reviewed so far, it appears that both object recognition and object categorization could benefit from framing their problems in a bidirectional framework that integrates cue diagnosticity and cue availability. The recognition example insisted on the impact task constraints and cue diagnosticity could exert on explanations of viewpoint-dependence in face recognition. The categorization example stressed the importance of the availability of perceptual cues in
explanations of simple categorization rules. We now turn to the nature of cue
diagnosticity and cue availability in everyday object recognition.

3. Everyday object recognition

What is everyday object recognition? Specifically, what are the information
requirements and the input information used when people categorize common
objects such as a car, a chair, a dog, and so forth? These issues are essentially
intertwined, but we will first explore the task demands of everyday object recogni-
tion before discussing input information.

3.1. The task demands of ‘everyday object recognition’

Classic categorization research has shown that the interactions between the
human perceiver and the objects of his or her world specify several hierarchical
levels of categorization. Following Rosch et al.’s seminal research, three of these
levels are often isolated: the superordinate (animal, vehicle, furniture), the basic
(dog, car, chair), and the subordinate (collie, Porsche, Chippendale chair) (Rosch et
al., 1976). Although these categorizations are all important, Rosch et al. showed that
one of them had a privileged status. When subjects were asked to spontaneously
name pictures of common objects, Rosch demonstrated that they preferentially used
basic-level names (see also Jolicoeur et al., 1984). Similarly, when asked to verify
that a picture belonged to a particular category, subjects’ decisions were faster for
the basic-level (Rosch et al., 1976). Together, these findings suggested that the
initial contact between the object percept and its semantic information occurs at
the basic level, also known in object recognition research as primal access (Bieder-
man, 1987), or entry point (Jolicoeur et al., 1984).²

It is usually thought that supplementary visual processing is required for subor-
dinate categorizations. For example, Jolicoeur et al. (1984) asked subjects to cate-
gorize common object pictures at their superordinate (animal), basic (dog) and
subordinate (collie) levels. In one condition (long exposure), pictures were presented
for 250 ms, in the other condition (short exposure) pictures were presented for 75
ms. Their finding was that while basic level categorizations were equally fast at long
and short exposures, short exposures disrupted the perceptually more taxing sub-
ordinate level categorizations.

The superordinate level is markedly different from the basic and subordinate
levels. For example, Rosch et al. (1976) determined that the shapes of objects within
basic and subordinate categories were generally more similar than those of objects
within superordinate categories (e.g. animal, vehicle, healthy food). Functional, not
so much perceptual, attributes tend to define superordinate categories. This explains
why Rosch and her colleagues found that basic and subordinate names facilitated
identification, at least more so than superordinate names. The functional attributes

²There are important differences between these concepts that will be discussed in Section 5.1.
associated with superordinate names would not allow the construction of a visual image that could facilitate the subsequent identification of a category exemplar.

For the reasons discussed so far, a significant portion of object recognition research has focused on faster basic level categorizations, and when it has not done so this has been considered a shortcoming. If common recognition tasks mirror basic level categorizations, then it is crucial to understand the organization of object categories at this level, to grasp the information requirements of the recognition task, and to explain why these elicit faster recognition. The basic level is often pictured as the most inclusive level at which objects ‘look alike’ in terms of their shape. One determinant of shape is part structure: objects with common parts tend to have a common shape. Tanaka and Gauthier (1998) define parts as ‘divisible, local components of an object that are segmented at points of discontinuity, are perceptually salient, and are usually identified by a linguistic label’ (p. 5). For example, the attributes ‘legs’, and ‘seat’ are the parts of the basic-level category chair.

Tversky and Hemenway (1984) found a dramatic increase in the number of parts listed from the superordinate to the basic level; non-part cues increased from the basic to the subordinate level, but little increase was found for parts. Thus, Tversky and Hemenway (1984) suggested that ‘the natural breaks among basic-level categories are between clusters of parts’ (p. 186). This claim is the basis of Biederman’s influential recognition-by-components (RBC) theory, which represents basic-level categories with the parts of their objects – specifically, with different geon structural descriptions (Biederman, 1987; Biederman and Gerhardstein, 1993). Hence, the widely-held assumption in the recognition literature that shape, as is represented by parts, is the most important information used in everyday, basic-level categorizations. Categories subordinate to the basic level are distinguishable on the basis of transformed parts and surface properties (Tversky and Hemenway, 1984). For example, the attribute legs and seat distinguishes chair from its contrast categories while long legs is a transformed part which distinguishes the subordinate category stool from its contrast categories. Thus, goes the argument, a qualitative identification of parts at the basic level must precede metrical computations between transformed parts at the subordinate level; the basic level is the entry point to the recognition hierarchy.

3.1.1. Difficulties with part information for all basic-level tasks

Two issues should be distinguished in explanations of the basic-level phenomenon. (1) The information requirements of basic level tasks: are parts really necessary and sufficient to distinguish between all memorized basic-level categories? (2) The available object information: is perception so organized that parts are the most salient object cues? On the one hand, if parts define the information requirements of everyday recognition, perceivers could have evolved to become primarily attuned to these object cues. On the other hand, if parts were not strictly necessary to distinguish between basic-level categories, a perceptual primacy for this information would be harder to justify in visual development.

Note that the issue here is not whether shape differences are relevant for basic level categorizations, but whether one representation of these shape differences –
part structure – is necessarily tied to the basic level. There is evidence that a ‘part-centric’ account of the basic level could be misguided. For example, Murphy (1991) demonstrated that cues other than parts could determine basic-level performance. Furthermore, findings in expert categorizations (Tanaka and Taylor, 1991) have revealed that the basic level was neither absolute nor unimodally specified, but could instead fluctuate with category expertise. Finally, available evidence that parts are the point of contact for basic-level categories is either based on feature listings, or reaction times. Without a conclusive argument for a necessary association between parts and the basic-level, evidence for parts is compatible with alternative shape representations that comprise part structures – e.g. silhouette, 2D edge configurations, or representations of the image at a coarser descriptive level. The following sections review each of these arguments in turn.

3.1.2. A basic level without parts

Since Rosch’s seminal research, an important issue has always been the extent to which the basic-level originates in the abstract organization of categories, or in the concrete constraints of perception. Murphy (1991), see also Murphy and Brownell (1985) suggested that the basic level was a consequence of the informativeness and the distinctiveness of a category representation in memory. Informative representations have many concrete object features; a representation is distinctive when it differs from contrast representations. In general, more specific representations tend to be more informative, but they are also less distinctive from other representations (Murphy, 1991). Thus, subordinate categories tend to score high on informativeness (e.g. two brands of cars convey detailed information), but low on distinctiveness (e.g. two brands of car are similar in overall appearance, at least more so than a brand of car and a type of shoe). In contrast, superordinate categories score low on informativeness, but high on distinctiveness (e.g. vehicle and furniture have different functions, shapes, parts, colors, textures, and so forth). On this account, the basic level would be a compromise between the accuracy of categorization at a maximally general level and the predictive power of a maximally specific level (Murphy and Lassaline, 1998).

Informativeness and distinctiveness are two constraints that apply to the abstract organization of categories in memory. Any type of object cue (including function, shape, color and textural cues) could functionally serve to optimize the informativeness and distinctiveness of categories, independently of its perceptual availability. Thus, the prediction is that parts should only specify the entry point when they optimize the informativeness and distinctiveness of the organization of the categories in memory.

In a series of experiments with artificial stimuli, Murphy (1991) questioned the necessity of parts for the basic level. His reasoning was that the addition of non-part information (here, mainly color and textural cues) to a basic level already structured with parts should not speed up its recognition advantage, if parts were the sole determinant of performance. Alternatively, if the informativeness and the distinctiveness of categories determined basic-level performance, then the addition of other cues to this level should enhance its informativeness and distinctiveness, and con-
sequently speed up its identification time. Results revealed that the additional colors and textures enhanced the basic-level advantage. Furthermore, categorization times increased with the addition of new cues at the subordinate level (because these categories were now less distinctive, sharing similar color and texture across exemplars) while superordinate categorizations were now faster (because they were more distinctive). A separate experiment showed that massing non-part information (color, size and texture) at the superordinate level eliminated the advantage of a basic level defined by parts: The diagnosticity of non-part cues at one level suppressed the diagnosticity of part information at the basic level. These results led Murphy (1991) to conclude that ‘... parts are neither necessary, nor sufficient to establish a basic-level structure... categorization into basic categories uses all kinds of information, not just part-based information’ (p. 436).

Although these conclusions could have profound implications for object recognition, there is an important rider. The experiments Murphy (1991) reviewed only demonstrate that a basic-level effect can be obtained with other cues than parts. Although the effect contrasts with the standard assumption that parts are necessary, it was obtained with artificial 2D stimuli whose part structures are different from those of typical 3D objects and might not tap into the same perceptual processes. This exemplifies the shortcoming that was leveled at categorization explanations that do not account for perceptual constraints. A differential availability of object cues might very well offer a speed advantage for real-world parts, or more generally shape cues, irrespective of the memory organization of categories, and so categorization theories must integrate this factor in their explanation of performance; distinctiveness should also concern perception, not just memory organization.

Bearing in mind that part descriptions are only one form of shape representation, a basic level effect could in principle also be obtained with alternative shape descriptions that comprise part representations (e.g. silhouettes and coarse-scale edge descriptions, fine-scale contours, depth maps and so forth). It should be an important goal of object recognition studies to tease out the relative contributions of different aspects of shape to basic level performance. Thus, it still is an empirical challenge to determine the information demands that structure basic level categories and the entry point to everyday recognition. The following discussion will suggest that these demands might not be the same for all object categories, but could instead change with category expertise.

3.1.3. A relative basic-level

If perceptual organization was such that parts from a fixed set (e.g. Biederman’s geons) were the primary ‘search keys’ into object memory, the level at which parts differentiate categories would fix the entry point to recognition. Research in conceptual expertise has questioned this idea of uniform information demands at the basic level. Tanaka and Taylor (1991) showed that extensive expertise with a category enhanced the speed of access of its subordinate categories, which became at least as accessible as basic-level categories. In a category-verification task, their subjects (dog and bird experts) first heard a category label (superordinate, basic or subordinate) and were then asked to indicate whether a subsequently presented
picture was an exemplar of the labeled category. For expert categories (bird or dog), the subordinate and basic categorizations were equally fast. Interestingly, the authors also discovered that the subordinate categories of experts were associated with more cues than their novice categories. Note that these cues were not supplementary parts (beagles do not have an extra nose, ear or leg that German shepherds lack), but cues that increased the informativity and distinctiveness (and the accessibility) of the expert category at its subordinate level. As Tanaka and Taylor (1991) put it succinctly, the basic-level could be in the eye of the beholder.

Although the ‘basic-to-subordinate shift’ was only ascertained in limited domains, expertise with a category is a continuum and the acquired salience of a subordinate level might unveil a pervasive principle of knowledge reorganization. There is suggestive evidence that perceptual expertise changes the defining cues and the categorization speed of other important object classes. For example, face identification is often thought to be a clear-cut ‘subordinate’ categorization because the similar global shape of faces makes them perceptually more taxing to discern. However, we are all ‘natural’ face experts, even if the ‘other race effect’, in which people discriminate faces of their own race with greater facility than those of another race (Brigham, 1986) suggests that this expertise is probably limited to faces of our own race. To the human observer, different views of an individual tend to be more alike than the same views of different persons. For faces of our own race, the entry point could therefore be the level of the individual (the level at which face views are more alike) instead of their assumed, ‘face’ basic level categorization (the level at which faces views differ more). Note that this could be reversed with faces of another race. Less familiar faces could appear comparatively more similar when seen from the same viewpoint. This illustrates another domain where a basic-to-subordinate shift might account for categorization differences between novices and experts.

One question that arises is the nature of the perceptual learning that would accompany such basic-to-subordinate shift. It has often been suggested that face expertise involves holistic recognition strategies (e.g. Carey, 1992). Tanaka and Farah (1993), see also Tanaka and Sengco (1997) showed that the identification of face parts was optimal when they were presented in their original configuration. Performance was impaired when parts were included in transformed configurations, and worse when they were presented in isolation. This suggests ‘holistic’, rather than independent representations of face parts. Recently, Gauthier and Tarr (1997) demonstrated that expertise with ‘Greebles’ (3D, computer synthesized ‘Martian beings’ designed to have the same kind of geometric homogeneity as faces) increased experts’ sensitivity to their configural cues. The principles underlying such expertise acquisition are reminiscent of those presented in Example 2 (see

3It is worth noting that the usage of ‘subordinate’ and ‘basic-level’ in object recognition and object categorization research do not always correspond. Object recognition researchers tend to use subordinate when the shapes of categories are alike, and basic-level when the shapes differ. Similarly, they tend to use subordinate for the identification of an individual face or object. In categorization, a subordinate level is the level immediately underneath the basic level (e.g. Ford Mustang is subordinate to car), which is first identified.
also Schyns and Murphy, 1994) in which new configural cues were created to distinguish between objects that initially look alike. The basic-to-subordinate shift could originate in the contribution of acquired configural cues to the informativeness and distinctiveness of expert subordinate categories. If the informativeness and distinctiveness of categories in memory determine the basic level, then the information requirements at this level might not be uniform for all categories, but instead depend on the individual’s level of expertise with them.

3.2. Towards a formal model of task constraints

In an attempt to formalize the notion of task constraints, Gosselin and Schyns (1997) developed a model that accounts for the basic-to-subordinate shift and other basic level phenomena. Their model, termed SLIP (strategy length information proxility), builds on Murphy and Brownell’s (Murphy and Brownell, 1985) idea that the ‘basic-levelness’ of a category originates in its informativeness and distinctiveness in a hierarchy of categories. The model assumes that categorization is a succession of tests on object cues such as ‘is X blue?’ and so forth. Some of these tests are highly informative and almost sufficient to isolate a category (e.g., ‘does X possess feathers?’ individuates birds) whereas others overlap between categories (e.g., ‘does X possess two wings?’ would be positive for birds and planes). SLIP derives its two main computational factors from this observation. To illustrate the first factor, consider the category hierarchy of Murphy and Smith (1982; Experiment 1, see also Murphy, 1991; Tanaka and Taylor, 1991) shown on the top structure of Fig. 4. Underneath the category names (e.g. hob, bot, com), the letters (e.g. a, b, c, d) designate the object cues that define three distinct category levels (the point preceding the cues corresponds to the cues inherited from the level(s) above the considered level – e.g. at the lowest level, a com is the feature conjunction acdeo). A careful observation of the structure should reveal that a single feature test is sufficient to identify each category, irrespective of its level. For example, the presence of o isolates com. Of course, several features often define categories, and in fact, three of them define mid-level categories in the top structure of Fig. 4. However, testing more than one feature does not add any information to the diagnosis; object cues are redundant for this categorization. The number of redundant cues at any category level in a hierarchy is the first computational factor of SLIP (it is three for the mid-level, and one for the other two levels).

In the real-world, however, features tend to overlap between categories. For example, to identify a helicopter, one needs to perform at least two tests: ‘does X fly?’ and ‘does X possess a rotor?’ Two tests are also necessary to identify a plane: ‘does X fly?’ and ‘does X possess two wings?’ Feature overlap in a category structure is the second computational factor of SLIP. It determines the minimal number of tests to arrive at a decision. To illustrate, consider the bottom structure of Fig. 4. To arrive at a com categorization, three tests (not just one, as in the top structure) must be executed: ‘does X possess a?’ ‘does X possess c?’ and ‘does X possess e?’ Two tests are necessary for bot, and only one for hob.
Category attentional slip integrates these two factors in a probabilistic model which performs the minimal number of tests to arrive at a category decision. High feature overlap between categories tend to increase their identification times. The model’s attention, however, sometimes slips from the ideal series of feature tests to a randomly chosen object cue. All things being equal, categories with high feature redundancy are more likely to benefit from attentional slip to a randomly chosen feature.

In a series of numerical simulations (see Gosselin and Schyns, 1997), the model predicted 20 out of 24 results from eight classic basic level experiments (including Murphy and Smith, 1982; (experiments 1 and 3) Murphy, 1991; (experiments 3 and 4 simple and enhanced, 5) Tanaka and Taylor, 1991; (Experiment 3, the basic-to-subordinate shift)). One established model of basic-level performance (Corter and Gluck, 1992; Category Utility) only predicted 12 of these results but that of Jones (1983), Category Feature Possession also predicted 20 results. It must be stressed that classic basic-level experiments only used categories organized with non-overlapping features. In other words, strategies of length 1 could be deployed and feature redundancy was the only factor affecting performance. However, the length of
strategies which results from feature redundancy should also affect basic-levelness, but this has so far been neglected in experiments on the basic level.

SLIP offers a powerful platform to explore the issues of basic level categorization and task constraints discussed so far. Starting from an objectively defined category structure (with its associated cue redundancy and cue overlap at different levels), an ideal categorizer can be constructed that predicts the relative accessibility (and speed of access) of each level in the hierarchy. Performance of the ideal is then compared to human subjects who categorize perceptually plausible experimental materials (e.g. computer-synthesized 3D objects composed of different parts, colors and textures). The comparative strategy enables us to explore how basic-levelness results from the interactions between objectively defined information demands and their relative perceptual availabilities. For example, if the presence of parts is the main determinant of faster recognition, then a basic level effect could occur at any level of the two hierarchies of Fig. 4. However, systematic variations of the objective information demands at each level (in the form of different shape, color or textural cues) could elicit performance discrepancies between the ideal and human subjects, revealing a contrast between objectively defined information demands and the perceptual limitations on their availability. This strategy could be applied to the understanding of the perceptual conditions of a basic-to-subordinate shift. Features created to solve expert categorizations must necessarily be diagnostic of these categories, have little overlap with other categories and have sufficient perceptual salience. Thus, newly-created features tend to augment feature redundancy at the subordinate level (the first component of category attentional slip) and minimize feature overlap and the length of categorization strategies at this level (the second component of the model). Such comparisons between an ideal categorizer (with its clearly-specified task constraints) and human performance might be one way forward in understanding basic level phenomena.

3.2.1. Lessons from task constraints

Everyday object recognition probably occurs at the basic-level and it is therefore important to understand the information demands it imposes on recognition tasks. The reviewed literature suggests that the basic level might not be designated by a unique, necessary and sufficient criterion (demands for parts) that determines the entry point to recognition. In contrast, there is suggestive evidence (1) that the basic level is the optimal level of informativity and distinctiveness of a category, (2) that parts are neither necessary, nor sufficient to determine basic level performance, but that other cues (e.g. color, texture, size, other shape cues such as shape configurations, an object silhouette and so forth) can also determine entry level, and (3) that perceptual experience with a category could change its defining cues and its basic-levelness. In sum, the information demands of a basic task could be relative and partially dependent on the individual’s experience (i.e. history of categorization) with an object category. Category attentional slip was proposed as a platform with which to oppose objectively defined information demands and their perceptual availability. Section 5 discusses implications of flexible basic level demands for recognition and categorization theories.
3.3. Perceptual information in object everyday recognition

Although task constraints depend on the functional organization of categories in memory and its associated information requirements, it is also important to understand how perception facilitates, or taxes, the extraction of relevant information. We know that object information is ultimately bounded by the retinal output (a million-dimensional space), but neuroanatomical and computational data suggest that this space is gradually projected (recoded) onto a much smaller-dimensional space of object cues. Unfortunately, too little is known about this dimensionality reduction process and the perceptual constraints it imposes on classification processes. Instead of attempting the impossible task of listing all likely information sources, this section examines the reduced set of object cues that determine performance in leading object recognition theories. We then relate these cues to the information demands of the categorization tasks they subsume, and discuss how task constraints and object information interact to determine performance.

Current thinking in recognition assumes that perception delivers shape cues to match against spatio-visual object representations in memory. To illustrate, in the RBC theory of Biederman (1987) the assumption that parts underlie the demands of basic-level recognition justifies the gradual projection of the 2D retinal input onto a space of geon structural descriptions (GSD, Biederman and Gerhardstein, 1993). The reduced object description is obtained in two stages. The first stage computes the edges of the input image. The second stage seeks 2D, non-accidental, viewpoint-invariant properties of edge descriptions (e.g. colinearity, curvilinearity, symmetry, parallelism, coterminal, see Lowe, 1987; Kanade, 1981). Collections of viewpoint-invariant labeled edges individuate the geons that compose the input object.

Together, these mechanisms should implement viewpoint-invariant recognition at the basic-level. More precisely, RBC predicts viewpoint-invariant performance whenever (1) the input object is decomposable into geons, (2) different GSDs represent different objects in memory, and (3) the same GSD is recoverable from different viewpoints (Biederman and Gerhardstein, 1993). Note that explanations of recognition performance that involve geons will, by construction of RBC, necessarily overlap with explanations that involve the viewpoint-invariant properties that individuate geons. Strictly speaking, it is these viewpoint-invariant properties of edges that determine the viewpoint-invariance of recognition.

In opposition to RBC, the view-based approach predicts viewpoint-dependent performance to stored object views. In attempting to define more precisely what constitutes a ‘view’, Tarr and Kriegman (1998) began to explore how viewpoint-dependent shape cues could determine performance. As an observer changes its vantage point, drastic changes often occur in the qualitative appearance of an object. For a given geometrical object class (e.g. smooth objects), a vocabulary of viewpoint-dependent labeled edges (local and multi-local edge configurations, see Tarr and Kriegman, 1998) can describe these qualitative changes. These cues partition the viewpoint space of an object into stable regions (in fact, into the views of an aspect graph of this object, Koenderink and Van Doorn, 1982). Viewpoint-depen-
dent recognition performance could partially result from enhanced perceptual sensitivity to viewpoint-dependent shape cues. Tarr and Kriegman’s psychophysical experiments demonstrated that this was the case, at least for some of the visual cues of their theory (Tarr and Kriegman, 1998).

From the above discussion, it appears that although RBC and the view-based approach both use shape information for recognition, the theories differ radically on the specific shape cues that they use. RBC suggests that perception detects 2D, viewpoint-independent cues that serve to reconstruct 3D geons, but Tarr and Kriegman showed that humans were sensitive to 2D, viewpoint-dependent edge configurations. Consequently, RBC and the view-based approach should in principle predict different recognition performance. In what follows, I will interpret performance as the interaction between the availability of cues and the information demands of the task at hand.

The image formation process entirely determines the availability of viewpoint-dependent and viewpoint-independent cues. The image formation process concerns the 2D retinal projection of the 3D geometry of an input object. It is likely that the 2D projection of a given 3D object comprises viewpoint-dependent and viewpoint-independent cues. Both types of cues should then be available to object recognition mechanisms, independently of whether or not a recognition theory advocates viewpoint-dependent (e.g. view-based) or viewpoint-independent (e.g. geon-based) representations of objects in memory. With this in mind, the explanation of performance shifts from the format of object representations in memory to a task analysis. Diagnostic recognition suggests that specific image cues are used when they become diagnostic for a particular task. What could be the object categorizations that would systematically require viewpoint-dependent versus viewpoint-independent object cues? Object recognition theories suggest a generic answer: basic-level categorizations would generally require viewpoint-independent image cues, and subordinate categorizations would demand viewpoint-dependent cues (e.g. Biederman, 1987; Jolicoeur, 1990; Tarr and Pinker, 1991; Farah, 1992). There are several difficulties with this claim.

A first difficulty arises from the recent results of Tarr and his colleagues (see Tarr et al., 1997) which revealed that even diagnostic combinations of shaded parts elicit viewpoint-dependent performance. Their experiments involved either a sequential matching or a naming of artificial, geon-based objects which varied in viewpoint between learning and testing. The objects comprised five parts, one, three or five of which were diagnostic geons, the remaining parts being non-diagnostic tubes. All objects were computer-synthesized and shaded. Results revealed a viewpoint-dependent performance in conditions in which Biederman and Gerhardstein’s criteria would predict viewpoint invariance (Biederman and Gerhardstein, 1993). This

*RBC predicts viewpoint-dependent performance in many practical situations of recognition. Rotation in depth of many real-world objects is such that different views will often convey different information (parts included, think, e.g. of a human body rotating in depth). Consequently, Biederman and Gerhardstein’s condition number three for viewpoint-independent performance (that the same part structural description is recovered from different viewpoints, Biederman and Gerhardstein, 1993) will not always be met, and both RBC and view-based theories will predict viewpoint-dependence.
questioned the overall viewpoint invariance of realistically-rendered geons, and the viewpoint-invariance of RBC (see also Edelman, 1995; Liu, 1996).

A second problem arises from evidence of basic-to-subordinate shifts with category experience. If these shifts occur together with the acquisition of new cues that change the informativeness and distinctiveness of categories, then it becomes difficult to systematically associate viewpoint-dependent or viewpoint-independent performance with a generic categorization level (either basic or subordinate). Instead, the issue really becomes whether the acquired cues that support faster categorizations at this level are viewpoint-dependent or not. As explained earlier, this depends on their projection on the 2D retina. The conclusion, then, is that the relationship between categorization levels and viewpoint-dependence should be treated with greater caution than is often the case in object recognition.

In sum, this section reviewed the object information assumed in major object recognition theories, and how it determines recognition performance. The analysis suggests (1) that the image formation process, not the format of object representations in memory, determines the availability of viewpoint-dependent and viewpoint-independent cues in the retinal input, (2) that dependence on viewpoint might therefore be better explained as an interaction between the cues requested for a given categorization and their dependence or independence on viewpoint in the input (see also Edelman, 1995; Liu, 1996; Tarr and Kriegman, 1998; for a similar view), and (3) that a systematic association between basic level information demands with viewpoint invariance and subordinate information demands with viewpoint-dependence might be misguided.

4. Interactions of spatial scales and diagnostic cues in scene recognition

The second aim of the paper concerns the psychological implications of the interactions between cue diagnosticity and cue availability. This section illustrates that the perception of a scene can change with the diagnosticity of its cues. Although a scene is not an object, but many objects, the example will raise issues that can directly be applied to studies of object recognition and categorization.

Computational vision and psychophysics have often emphasized the importance of simultaneously processing stimuli at multiple spatial resolutions, called ‘spatial scales’ (Campbell and Robson, 1968; Blackmore and Campbell, 1969; Breitmeyer and Ganz, 1976; Marr, 1982; Burt and Adelson, 1983; Canny, 1986; Ginsburg, 1986; Mallet, 1989; de Valois and de Valois, 1990; among many others). Starting with the observation that recognition algorithms could hardly operate on the raw pixel values of digitized images, vision researchers investigated multi-scale representations to organize and to simplify the description of events. Coarse-to-fine processing suggests that it may be computationally more efficient to derive a coarse and global (but imprecise) skeleton of the image before fleshing it out with fine grain (but considerably noisier) object cues (Marr, 1982; Watt, 1987). Evidence of coarse-to-fine processing in human psychophysics has been reported for face (e.g. Sergent,
Fig. 5 illustrates the perceptual scales that are available to recognition mechanisms (adapted from Schyns and Oliva, 1994). Reconstruction of the image from fine scale edges (technically, the high spatial frequencies, HSF) should unveil a city scene in the top picture and a highway in the bottom picture. However, if this information is made unavailable (either by squinting or blinking while looking at the pictures) your perception of the top picture should turn into a highway and a city in the bottom picture (step back from Fig. 5 if this demonstration does not work). You would then perceive the coarse scale of the pictures (technically, the low spatial frequencies, LSF).

Even though there is now little doubt that the visual system operates at multiple scales, their selection for recognition is still a matter of current debate. According to psychophysical evidence, the earlier availability of coarse scales should force a coarse-to-fine recognition scheme. However, this claim disregards the influence the task could exert on scale selection. In the diagnosticity framework, the information requirements of a task change the diagnosticity of the cues which are actively sought in the input. If diagnostic cues were preferentially associated with a different spatial scale, then the task at hand could determine scale selection. For example, whereas coarse scale information might be sufficient for a city categorization of a city picture, its New York categorization might necessitate comparatively finer scale cues (see Fig. 5). Such top-down influence of task constraints on scale perception is particularly interesting given the precedence the latter has on many processes of early vision such as stereopsis (Schor et al., 1984; Legge and Gu, 1989), motion (Morgan, 1992), depth perception (Marshall et al., 1996) and saccade programming (Findlay et al., 1993).

In their Experiment 2, Oliva and Schyns (1997) tested the prediction that the diagnosticity of scale cues could change the perception of identical hybrid scenes. Hybrids have the main advantage that they multiplex (i.e. combine) information across spatial scales. They can control which scale supports which categorization. To illustrate, a city categorization of the top picture of Fig. 5 would unambiguously attest the usage of coarse scale cues (LSF), whereas its highway categorization would indicate the usage of its fine scale cues (HSF). The issue was whether such distinct categorizations could simply originate from one scale of the picture (either coarse or fine) being more diagnostic than the other for the task at hand.

In a sensitization phase, two subject groups (the LSF and the HSF groups) were asked to categorize hybrids which presented a scene at only one spatial scale—the other scale represented meaningless patterns, see Fig. 6. Each hybrid was presented for 135 ms on a computer screen, and subjects were instructed to name aloud the scene they saw. It was expected that these stimuli would implicitly sensitize perceptual processes to tune to the informative scale (either LSF or HSF, depending on the group). In a testing phase immediately following sensitization, subjects were shown hybrids whose LSF and HSF represented two different scenes (as in Fig. 5). The sensitization and testing phases were appended so that subjects did not notice a transition, or a change of stimuli; from their viewpoint, they were categorizing a single series of pictures.
Categorization data revealed that subjects maintained their categorizations of ambiguous hybrids at the scale congruent with their sensitization. For example, hybrids were constructed by combining the low spatial frequency (LSF) components of one scene (e.g. a highway) with the high spatial frequency (HSF) components of another scene. The top picture mixes the LSF of a highway and the HSF of a city. The bottom picture mixes the LSF of a city with the HSF of a highway.

Fig. 5. This figure (adapted from Schyns and Oliva, 1994) shows examples of the hybrid stimuli. Hybrids were constructed by combining the low spatial frequency (LSF) components of one scene (e.g. a highway) with the high spatial frequency (HSF) components of another scene. The top picture mixes the LSF of a highway and the HSF of a city. The bottom picture mixes the LSF of a city with the HSF of a highway.
LSF subjects categorized the top hybrid of Fig. 5 as highway while HSF participants categorized the same picture as city. Orthogonal categorizations of identical pictures indicate that different cues were used, but they do not necessarily imply, however, that the stimuli were orthogonally perceived. Subjects might have perceived the two scenes composing the ambiguous hybrids but strategically decided to report only
one of them. Alternatively, the requirement of locating diagnostic cues for the task might have changed scale usage to the point of changing the perception of identical pictures. In a debriefing following the experiment, participants were told what hybrid stimuli were and were asked whether they saw any such stimulus during the experiment. Out of 24 subjects, only one reported seeing two scenes; all others were surprised to learn that they were exposed to ambiguous stimuli. Follow-up studies (see Oliva and Schyns, 1997) confirmed that neglected information at the non-diagnostic scale did not enter the scene percept, but was nonetheless registered at a lower level of processing.

In sum, the example illustrates that another recognition phenomenon, scale-based recognition, might be better explained in a bi-directional framework, as an interaction between the perceptual availability of multiple spatial scales and the requirement of locating diagnostic scale cues, rather than as a unidirectional, perceptually determined process. The example also shows that the information demands of a categorization can flexibly change the immediate appearance of an identical stimulus. Flexible perceptions could be the psychological byproduct of the interactions between task and cues. It is unclear whether task-driven perception generalizes from the experimental room to everyday situations of face, object and scene recognition. However, entertaining its possibility opens a number of issues that we address in Section 5.

5. General discussion

It was the main goal of this paper to establish a dialogue between object recognition and object categorization theories, with the intention of raising issues that could cross-fertilize their research. To this end, I presented diagnostic recognition, a framework which expresses object recognition and categorization phenomena as interactions between the information requirements of specific categorization tasks and the perceptual information available from visual inputs. Diagnostic recognition insists on the diagnosticity of perceptual object cues in a task to understand object recognition and categorization phenomena. Two examples illustrated the opposite benefits that object recognition and categorization theories could gain from explicitly considering the two factors of diagnostic recognition in their explanations of performance. The face example showed how object recognition could benefit from more extensive studies of task constraints, and the Martian cells categorizations suggested that the availability of perceptual cues might need further consideration in explanations of even very simple categorization problems.

The second part of the paper extrapolated the approach of diagnostic recognition to the account of ‘everyday recognition’ performance. Everyday recognition was equated with the basic level of a categorization hierarchy. Examination of information demands at the basic level from a categorization perspective suggested that it was the optimal level of informativity and distinctiveness of a category, that parts were neither necessary nor sufficient to structure this level, but that many other cues could elicit a basic level effect. It was also suggested that the individual’s perceptual
experience with an object category could change its defining perceptual cues, as well as its entry point. Category attentional slip was proposed as a platform to formalize task constraints and to explore how they fit, or contrast with the perceptual availability of object cues.

Turning to the object information of leading recognition theories, it was first observed that perception is often assumed to initially deliver shape cues to match into object memory. These are either viewpoint-dependent or viewpoint-invariant depending on the 3D geometry of the object and its projection on the 2D retina. The usage of one type of cue or another (and the viewpoint-dependent or viewpoint-independent nature of recognition performance) was then related to the information demands of the task at hand. The faster recognition of subordinate categories by experts suggests that information demands at the basic level will vary with the acquisition of category expertise. Hence, it is unlikely that the basic level (or the subordinate level) will be uniformly viewpoint-independent (or viewpoint-dependent).

The scene recognition example addressed the problem of multiple categorizations in the perceptually plausible materials of spatial scales. The behavioral data suggested that a bi-directional account including top-down task constraints provided a better explanation of the opposite categorizations of identical scene pictures than a unilateral, bottom-up flow of constraints. The example also illustrated a situation in which different information demands induced the flexible perceptions of complex visual stimuli, which suggests that the diagnosticity of cues not only changes their use in the input, but also the immediate appearance of the stimulus.

5.1. Implications of diagnostic recognition for studies of object recognition, categorization and perception

The idea of a flexible basic level raises a number of new issues for object recognition and categorization theories. Consider the example of an Asian person who has had sufficient experience with Caucasian faces that she can individuate them. Presumably, this person has learned the cues that enable precise discriminations within the homogeneous category of Caucasians (the example is of course reversible). We could expect face identity to stand out in informativeness and distinctiveness, and she should classify Caucasians the way Caucasians do, equally fast at their subordinate and basic levels (Tanaka and Gauthier, 1998).

Equal accessibility of categorization levels, however, does not necessarily imply replacement of one level by the other. One could legitimately ask whether, following even extensive expertise, the categorization sequence is still basic before subordinate (e.g. face before Sara, or car before Mustang), or whether the entry point has genuinely changed to the subordinate level (e.g. Sara and Mustang). This issue is difficult because too little is known about the nature of the perceptual changes that can accompany category expertise. If new diagnostic features are simply added at the subordinate level, then it is difficult to assume that object information at the inclusive basic level would simply be dismissed. For example, even if Dalmatians have a diagnostic texture, they still have a typical dog shape, and the dog expert
would probably not confuse a Dalmatian dog with a Dalmatian texture wrapped around a car.

There are alternatives to an additive version of category expertise which might support independent access to the subordinate level. For instance, extensive experience with a category could involve non-linear feature-learning mechanisms. Subordinate features could then represent an altogether different encoding of the entire object – much in the way in which the xy configural encoding was not reducible to the additive perception of feature x and feature y in Example 2. In fact, the argument has been made (e.g. Diamond and Carey, 1986) and evidence exists (Schyns and Rodet, 1997; Gauthier and Tarr, 1997) that expertise with geometrically homogeneous objects enhances sensitivity to their configural cues – cues that are not perceived as a simple addition of their components. If expertise with a category involves such non-linear recodings of its objects, then independent access at the subordinate level would not necessarily lead to the information loss implied by an additive model. The relationships between independent entry levels and feature learning mechanisms clearly deserve further research.

These speculations about the nature of the entry point to recognition are important because they directly question the nature of ‘the right’ representation of an object. For example, the idea of a basic level that can progressively shift as people acquire new features that increase the informativeness and distinctiveness of subordinate categories implies that the most stringent task (e.g. face identity) in which an object is used determines its representation. Do we maintain in memory different representations of the same object for different categorization tasks, or do we acquire and progressively transform a unique perceptual representation that encodes the information demands of the most stringent subordinate tasks while preserving those of less taxing tasks (such as the former basic level)? If visual cognition constructs multiple representations, then independent entry points for recognition would mean that different tasks access different representation systems. By contrast, independent entry points to a unique representation would mean that certain information of this representation elicits faster recognition times. This dichotomy reflects the ambiguity existing between basic-levelness (which is a functional measure of recognition performance computed over a category organization) and entry point (which is a structural statement about the access to internal systems of object representations). The relationship between multiple categorizations of an object and its representation(s) should be carefully considered.

Intertwined with these issues are the relationships existing between the information demands of a categorization task (the diagnosticity of object cues) and the perceptual appearance of the visual input. The logic of diagnostic recognition implies that diagnostic cues must be clearly perceived in the visual input for the task to succeed. What happens, then, to the other, non-diagnostic cues? Is perception flexible and correlated with the information demands of a categorization task, or is it instead ‘veridical’ and independent of the task at hand? The scene example presented earlier illustrated one intriguing instance of a flexible, task-dependent influence on the perceptual appearance of stimuli. Do flexible percepts also arise when people categorize the same chair as a chair, or as an Art Deco chair, or a scene
picture as city or New York, or a face as face or Mary? In short, do perceptions follow, even in part, task constraints? It is difficult to derive predictions from existing recognition and categorization theories because most simply do not address the question. However, these issues are fundamental because they tap into the ‘cognitive penetrability’ (Fodor, 1983) of early visual processes and suggest that cue diagnosticity could partially determine object perception. I believe there is much to be gained from studying how categorization influences perception.

5.2. Limitations of diagnostic recognition for studies of object representations

Diagnostic recognition is a framework in which the information goals of object categorization tasks are considered before their perceptual representations.

Although this is a good, generally recommended approach to theory construction (e.g. Marr, 1982), it nevertheless presents serious limitations for the study of object representations.

The reason is simply that thinking from task constraints to their perceptual representations could over-represent the considered information demands in the proposed representation. For example, if it were discovered that the information requirements of an object categorization were X, then it would be an easy step to assume that the representation of this object was effectively that X. But then, how would we know whether X represents the object, or the task itself?

It is interesting to note that major recognition theories sit on the opposite sides of the spectrum of independence of representations from task demands. Biederman’s geons directly mirror the assumed information demands (parts) of the basic level (Biederman, 1987). The view-based approach stands on the other side of the spectrum. Unless more clearly specified, an object view could potentially represent all the information that can ever be requested from an object seen from this view, including parts. In other words, a view is too powerful a representation, and it is an important research goal to attempt to reduce the high dimensionality of a view to a low-dimensional subset of object cues (e.g. Edelman and Intrator, 1998; Tarr and Kriegman, 1998).

However, it still is the case that diagnostic recognition might be better suited to explain recognition performance and to study object perceptions than to infer representation formats in memory. Performance involves both object and input representations: The input image must first be encoded with object cues for matching against memorized representations. Because there is no general theory specifying the information content of the 2D projections of 3D objects, behavioral performance might not be sufficiently powerful to isolate issues of object formats in memory from issues of input information and perception (Liu, 1996). Performance that might be attributed to a particular object format might also be attributed to the interaction between task demands and the usage of specific image cues. For these reasons, diagnostic recognition suggests that object and input representations should be unified and constitute the set of image cues that are available and perceived for different object categorization tasks. Hence, it insists on flexible perceptions rather than various systems of object representations.
There is a similarity between the framework presented here and the ideal observer approach to recognition (Bennett et al., 1993; Liu et al., 1995). Both diagnostic recognition and the ideal observer stress that the available object information and the perceptual constraints on its extraction influence performance (see also Liu, 1996). However, actual developments of the ideal observer do not include task constraints which "act" on different object cues to assign them different diagnosticities. For example, Liu (1996) suggests that a viewpoint-independent framework for representation should predict that the performance only depends on the information content of the input image. However, there is suggestive evidence that recognition performance is also dependent on the task itself. Extensions of ideal observers might need to include the notion of flexible task constraints.

6. Concluding remarks

This paper started with a question: 'how could we frame issues common to object recognition and categorization?" A bi-directional framework was presented which expresses recognition and categorization phenomena as interactions between the information demands of categorization tasks and the perceptual availability of object information. This integrative framework raises new issues that could shed some light on the nature of face, object and scene perception.

Acknowledgements

The author wishes to thank Michael Burton, Simon Garrod, Frédéric Gosselin, Gregory Murphy, Martin Pickering and two anonymous reviewers for helpful comments on an earlier version of this manuscript.

References


