Psychonomic Bulletin & Review
2001, 8 (3), 454-469

Usage of spatial scales for the categorization of faces, objects, and scenes

DONALD J. MORRISON and PHILIPPE G. SCHYNS
University of Glasgow, Glasgow, Scotland

The role of spatial scales (or spatial frequencies) in the processing of faces, objects, and scenes has recently seen a surge of research activity. In this review, we will critically examine two main theories of scale usage. The fixed theory proposes that spatial scales are used in a fixed, perceptually determined order (coarse to fine). The flexible theory suggests instead that usage of spatial scales is flexible, depending on the requirements of visual information for the categorization task at hand. The implications of the theories are examined for face, object, and scene categorization, attention, perception, and representation.

In recent years, a number of papers exploring the role of spatial scales (or spatial frequencies) in the processing of natural images, such as objects, scenes and faces, have been published. This empirical work has employed a variety of methodologies to address a number of related issues. Given that activity in this research area has increased considerably of late, we believe now is the right time to take stock, critically review the work done so far, draw conclusions, and make some suggestions for future investigation. This is the objective of the present article.

Studying scale usage for the categorization of complex visual images is important if we are to understand how visual perceptual and cognitive processes operate, thus enabling us to interact efficiently with our complex visual environment. Spatial filtering is usually thought to be an early stage of visual processing, the outputs of which form a basis for higher level operations, such as categorization and recognition. A complete account of these higher level processes will therefore require a good understanding of early visual processes (e.g., spatial filtering) and the constraints they impose. But why focus on spatial filtering when there are other important dimensions of early vision, such as color and depth? Luminance variability in the visual field, arguably a crucial source of information for recognition, is encoded by spatial filters. For example, the encoding of detailed edges portraying the contours of a nose, eyelashes, the precise shape of the mouth and eyes, and so forth can be traced to spatial filters operating at a fine spatial resolution (i.e., high spatial frequencies; HSFs). In contrast, spatial filters at a coarser resolution (i.e., low spatial frequencies; LSFs) could encode pigmentation and shape from shading from the face. That is, spatial filters encode a wide range of useful visual information, at least those cues thought to be critical for everyday face, object, and scene categorization. Hence, specifying the nature of scale usage might have implications for how we categorize the sorts of complex visual inputs we encounter in everyday life.

Research on spatial filtering is an established tradition of psychophysics, but it does have implications for theories of recognition/categorization.[1] As we will explain, studying scale usage is an excellent medium for examining the interactions between perceptual and cognitive processes. For example, if visual cues used for different categorizations of an identical input (face, object, or scene) are associated with distinct spatial frequencies, low-level processing of spatial frequencies could constrain categorization. On the other hand, the categorization task could itself influence the output of early perceptual processes. At a more general level, the cognitive impenetrability of vision can be addressed (Fodor, 1983; Pylyshyn, 1999; Schyns, Goldstone, & Thibaut, 1998).

To begin, we introduce the key concepts of spatial scales and spatial frequency channels. Theories of scale usage are then described and evaluated in light of empirical findings. We discuss some methodological pitfalls inherent in this type of research and conclude by highlighting some issues we think are interesting and require further exploration.

SPATIAL SCALES AND SPATIAL FREQUENCY CHANNELS

Natural images in our environment provide us, the viewers, with a wide spectrum of spatial information, ranging from extremely coarse to very fine. Fine spatial information is associated with detailed parts of the image, whereas coarse spatial information corresponds to larger, less detailed parts. This spectrum of spatial information can be described using Fourier analysis (Campbell & Green, 1965; Davidson, 1968). Accordingly, the coarse spatial information in the image is referred to as the LSFs, and the fine spatial information is referred to as the

Correspondence concerning this article should be addressed to P. G. Schyns, Department of Psychology, University of Glasgow, 58 Hillhead Street, Glasgow, G12 8QB, Scotland (e-mail: philippe@ psy.gla.ac.uk).

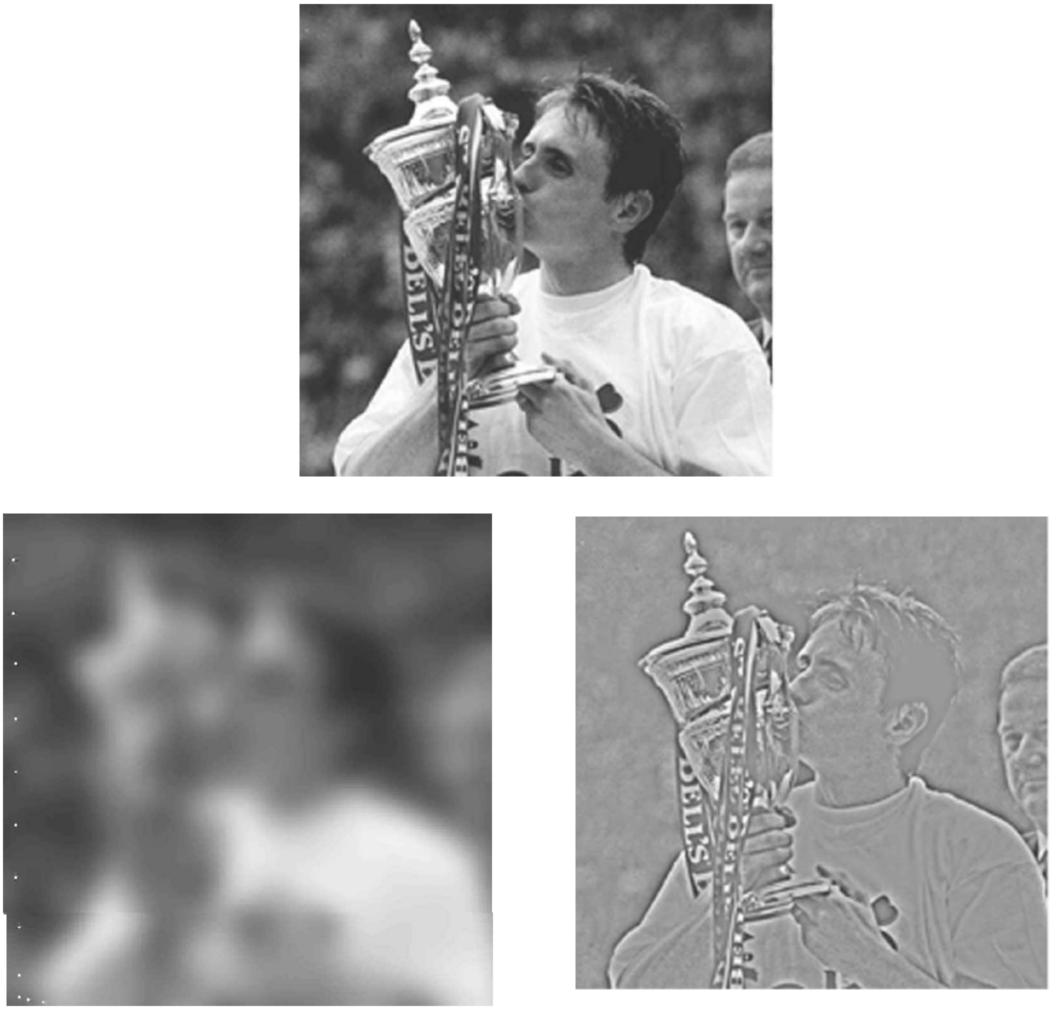

Figure 1. Examples of spatial scales. The full spectrum of spatial information is depicted in the original (top) image. The low spatial frequencies alone can be extracted from this image (bottom left), as can the high spatial frequencies (bottom right).

HSFs. Examples of LSFs and HSFs derived from a natural image are shown in Figure 1. The original image (top), containing the full spectrum of spatial information (i.e., all spatial frequencies), clearly depicts a sportsman kissing a trophy. Also shown are the LSFs (bottom left) and the HSFs (bottom right) derived from the original image. The LSFs can be seen to correspond to the coarse, less detailed parts of the scene: Such properties as the color and luminance of blobs are carried in the LSFs whereas fine details, such as the writing on the ribbons of the trophy, are discarded. Luminance blobs provide a useful skeleton of the image, from which fine details can be fleshed out. On the other hand, HSFs represent the more detailed aspects of the image: HSFs preserve fine details, such as the writing on the ribbons, but not coarse properties, including the color and luminance of blobs.

Spatial frequencies are typically defined as a number of cycles per degree of visual angle and/or a number of cycles per image. For example, the coarse-scale information depicted in Figure 1 (bottom left) are those spatial frequencies below a number of cycles per degree of visual angle when viewed from a given distance, or 8 cycles per image. The fine-scale information (bottom right) corresponds to those spatial frequencies above a number of cycles per degree of visual angle when seen from a given distance, or 24 cycles per image. Cycles per degree of visual angle is a relative measure, taking the viewing distance of the observer into account, whereas cycles per image is an absolute measure of information content, irrespective of the observer. The former terminology is ubiquitous to psychophysics, whereas the latter is more typical of image processing.[2]

A *spatial frequency channel* is a filtering mechanism that passes a restricted range of the information it receives. There are three types of spatial frequency channels. A low-pass channel passes all spatial frequencies below a particular cutoff, while discarding all frequencies above this cutoff. Conversely, a high-pass channel retains all frequencies above a cutoff, while discarding those below it. The original picture shown in Figure 1 was *low-passed* in order to extract only the spatial frequencies below 8 cycles per image (bottom left) and *high-passed* to derive those above 24 cycles per image (bottom right). The third type of spatial frequency channel is termed a *bandpass channel.* Such a filter passes only the frequencies between two cutoffs, discarding those at each end.

The ideas introduced above are important when considering how we process natural visual stimuli, because, as has been demonstrated by psychophysical studies, early vision filters the input with a number of channels, each tuned to a different bandwidth of spatial frequencies (see DeValois & DeValois, 1990, for an excellent review of spatial vision). For example, in their seminal paper, Campbell and Robson (1968) reported that the detection and discrimination of simple sinewave patterns was predicted by the contrast of their individual component spatial frequencies. This was only possible if the visual system was decomposing the patterns with spatial frequency filters, and so the authors concluded that early vision comprises groups of quasilinear (i.e., additive) bandpass filters, each tuned to a specific frequency band (see also Graham, 1980; Pantle & Sekuler, 1968; Thomas, 1970; Webster & DeValois, 1985). Frequency-specific adaptation studies showed that these channels could be selectively impaired in their sensitivity to contrast, suggesting that they are independent (e.g., Blackmore & Campbell, 1969).

It therefore appears that the visual input is initially processed at multiple spatial scales, functionally described by about four to six spatial frequency channels (Ginsburg, 1986; Wilson & Bergen, 1979). Subsequent developments indicated that these channels are interactive (e.g., Henning, Hertz, & Broadbent, 1975) and nonlinear (e.g., R. J. Snowden & Hammett, 1992). However, it is still generally agreed that spatial filtering occurs prior to other early forms of visual processing, including stereopsis (e.g., Legge & Gu, 1989), motion perception (Morgan, 1992), depth perception (Marshall, Burbeck, Ariely, Rolland, & Martin, 1996), and saccade programming (Findlay, Brogan, & Wenban-Smith, 1993). Spatial filters consequently provide an excellent candidate for the building blocks of visual perception that might determine visual categorizations.

THEORIES OF SCALE USAGE FOR CATEGORIZATION

Given that vision is equipped to filter the input at multiple spatial scales, an important question concerns how information from these channels is used to categorize the complex visual stimuli we encounter in everyday life. On the one hand, early constraints on the extraction and availability of coarse- and fine-scale information may impose a fixed order on their usage for categorization. More recently, however, it has been suggested that such a fixed view of scale usage may be misguided and that we should instead consider scale usage as flexible and dependent on the current task demands.

Fixed Usage: Coarse-to-Fine Hypothesis

A view commonly held by researchers in this area is that there is a fixed coarse-to-fine bias for scale processing, with respect to both the sensory processing of scale information and its usage for face, object, and scene categorizations (e.g., Breitmeyer, 1984; Fiorentini, Maffei, & Sandini, 1983; Parker & Costen, 1999; Parker, Lishman, & Hughes, 1992, 1997; Schyns & Oliva, 1994). The roots of this idea can be traced to classical research in physiology. Enroth-Cugell and Robson (1966) examined the spatiotemporal characteristics of X and Y retinal ganglion cells. They discovered that whereas X cells respond in a sustained way to high-resolution stimuli, Y cells respond more transiently to low-resolution stimuli. X and Y retinal ganglion cells were, therefore, differentiated on both their temporal (sustained vs. transient) and spatial (low and high resolution) properties. Hubel and Wiesel (1977) demonstrated that these distinctions were preserved at the lateral geniculate nucleus. Y cells dealing with a transient, gross analysis of the stimulus, project exclusively to the magnocellular layers of the lateral geniculate nucleus, whereas X cells, concerned with a sustained and detailed analysis, project to both parvo- and magnocellular layers. These spatiotemporal and anatomical distinctions influenced the early computational models of visual processes, including edge extraction, stereopsis, and motion (see Marr, 1982, for discussions and examples).

If physiology prompted the idea of a coarse-to-fine processing in early vision, researchers in higher level vision soon realized that a multiscale representation of the image could be used to organize and simplify the description of events (e.g., Marr, 1982; Marr & Hildreth, 1980; Marr & Poggio, 1979; Watt, 1987; Witkin, 1987). For example, edges at a fine spatial resolution are known to be noisy and to represent confusing details that would not be apparent at a coarser resolution. However, fine-scale details are often required to distinguish between similar objects or whenever the task requires detailed information. An effective processing strategy may therefore produce a stable description of the image before the noisier information is extracted. Accordingly, a stable, but less detailed, coarse description of the image would first be produced (see the bottom left image of Figure 1), which would then be fleshed out with the fine-scale information often required for successful categorization (see the bottom right image of Figure 1). That is, the LSFs may be extracted and used to recognize stimuli before the HSFs. We call this the *coarse-to-fine hypothesis.*

However, the notion of a coarse-to-fine recognition strategy is often assumed but rarely explicitly stated. The general view is that "the lower spatial frequencies in an image are processed relatively quickly while progressively finer spatial information is processed more slowly" (Parker & Costen, 1999, p. 118). The precise status of the coarse-to-fine hypothesis is therefore unclear. Is the physiological bias in the temporal availability of coarse- and fine-scale information (with LSFs being extracted before HSFs) so constraining as to result in a coarse-to-fine strategy of using scale information for categorization (i.e., a perceptually driven coarse-to-fine categorization scheme)? Or is there a coarse-to-fine categorization strategy for a quite separate reason—namely, that an efficient scheme for the recognition of complex images first produces a coarse skeleton, which is then fleshed out with fine-scale details (i.e., a strategically driven coarse-to-fine categorization scheme)? If the latter, the direction of scale processing may be manipulable by task demands, rather than fixed in physiology.

Marr's (1982; Marr & Hildreth, 1980) theory of the primal sketch has had an important influence on theories of recognition. A coarse-to-fine bias exists at the first stage of description of the input (the primal sketch). Stable coarse scale information is used before progressing to the less reliable fine-scale information when locating intensity changes (or zero crossings) in the image. Watt (1987) also suggested that when a stimulus remains in view, the range of the spatial filter shrinks in size over time—that is, spatial filters start to operate at a coarse scale, before shrinking to operate at progressively finer scales. However, construction of the raw primal sketch is among the earliest stages of processing in Marr's model, and such a bias in scale processing may or may not extend to categorization processes. Nevertheless, the view that there is a coarse-to-fine bias in the usage of spatial scales for recognition has permeated this research area (e.g., Breitmeyer, 1984; Fiorentini et al., 1983; Parker & Costen, 1999; Parker et al., 1992, 1997; Schyns & Oliva, 1994). Accordingly, the first theory of scale usage proposes that the most effective route to recognition would be via coarse-scale information, which is subsequently fleshed out with higher spatial frequencies (e.g., Schyns & Oliva, 1994; Sergent, 1982, 1986). The perceptual versus strategical status of this coarse-to-fine scheme was not addressed until recently.

Flexible Usage Hypothesis

An alternative to the fixed coarse-to-fine hypothesis of scale usage for categorization has recently been put forward by Oliva and Schyns (1997; Schyns & Oliva, 1999). Consider the full bandpass image depicted in Figure 1. This image may be categorized in a number of ways—for example, as a person, as a male, or to those who know him, as Tom Boyd, captain of the Glasgow Celtic Football Club. Furthermore, distinct categorizations of this image will require different perceptual cues, which themselves could be associated with different regions of the spatial spectrum. For example, Schyns and Oliva (1999) showed that the perceptual cues most useful for determining the identity, gender, and expression of a face may be associated with different spatial resolutions (see also Sergent, 1986). Therefore, when categorizing an image, there may be a bias in favor of the spatial scales with which task-relevant perceptual cues are associated. Rather than being fixed (coarse to fine), Schyns and Oliva (1999) suggested that scale usage for categorization may be flexible and determined by the usefulness (or diagnosticity) of cues at specific scales. We refer to this as the *flexible usage hypothesis.* Unlike this view, the coarse-to-fine hypothesis neglects the nature of the categorization task and its information requirements. An interesting consequence of flexible scale usage is that the perceptual processing of an identical visual input may be influenced by the nature of the categorization task (Schyns, 1998). Indeed, evidence does suggest that higher level processing can influence the construction of an image percept (e.g., Schyns et al., 1998). We now describe and evaluate the empirical work addressing the fixed versus flexible usage of spatial scales in the processing of visual stimuli.

EMPIRICAL INVESTIGATIONS OF SCALE PROCESSING

A Coarse-to-Fine Bias for Scale Extraction?

Results from a number of psychophysical studies that show that processing times of sinusoidal gratings are influenced by spatial frequency suggest that there is a coarse-to-fine bias for the extraction of spatial information. The time taken to detect the onset, offset, or contrast reversal of a sinusoidal grating increases approximately monotonically with spatial frequency, even when contrast of the gratings is equated (e.g., Breitmeyer, 1975; Gish, Shulman, Sheehy, & Leibowitz, 1986). For example, observers in Parker's study saw vertical sinusoidal gratings ranging from 1 to 12 cycles per degree of visual angle and responded when a stimulus appeared, or disappeared or there was a 180° phase shift. Response latencies for all three conditions increased with spatial frequency. A similar delay in processing higher spatial frequencies has been found by recording visually evoked responses (e.g., Mihaylova, Stomonyakov, & Vassilev, 1999; Parker & Salzen, 1977) and using a perceptual matching task. Psychophysical work with simple sinewave gratings seems to indicate that coarse-scale information (e.g., contrasts at different orientations) becomes perceptually available before fine-scale information.

However, such findings do not imply the existence of a similar bias for the recognition of natural images. The relationship between spatial frequency and reaction time was found by using simple stimuli where patterns were single frequencies presented at one orientation. Complex pictures have energy at multiple spatial frequencies and orientations, and the patterns these represent could induce different perceptions of identical contrasts at different locations of the image. Thus, any bias found with simple sinewave gratings might not transfer to more complex patterns. A further and perhaps more important issue is whether a low-level perceptual bias would be so constraining that it would impose a mandatory coarse-to-fine recognition scheme. The time course of scale availability may have little or no influence on the scale initially used for recognition. In other words, early biases in scale perception might not necessarily translate into the same biases in scale-based recognition (Oliva & Schyns, 1997). We now turn to the issue of scale usage for the categorization of complex pictures.

A Coarse-to-Fine Categorization Strategy?

Since the psychophysical studies described above, experiments exploring scale usage in the processing of complex visual stimuli, such as faces, objects, and scenes, have been reported. Parker et al. (1992) were the first to provide evidence of a coarse-to-fine bias in scene processing. The subjects in this study (Experiments 2 and 3) rated the perceived picture quality of short (120-msec) sequences comprising three images presented for 40 msec each without interval. High- and low-passed versions of one scene were used, in addition to the original full-spectrum picture. The order in which the spatial content was presented within a sequence was manipulated—that is, sequences were either low-to-high or high-to-low. Observers rated low-to-high sequences as being of better quality than high-to-low sequences comprising exactly the same images. The subjects were also more likely to indicate the presence of the full bandwidth image in low-to-high than in high-to-low sequences, regardless of whether or not this stimulus was included. In a subsequent study (Parker et al., 1997), observers were again shown 120-msec sequences comprising full bandwidth and/or degraded images and were required to discriminate sequences that contained filtered images from those that did not. Low-to-high sequences were more likely to be mistaken for full bandwidth presentations than were high-to-low sequences, whether the images were derived from the picture of a scene (Experiment 1) or from the picture of a face (Experiment 2). Parker et al. (1992, 1997) suggested that spatial information in a coarse-to-fine sequence is integrated more efficiently than that in a fine-to-coarse sequence.

However, an argument we put forward when considering the psychophysical studies looking at scale availability is also appropriate here, since it remains unclear whether a coarse-to-fine bias for the perceptual *integration* of spatial scales would necessitate a similar bias for recognition. A further difficulty with these studies is that the interesting technique of presenting a series of spatial frequency information is associated with the unfortunate side effect of backward masking, whereby a particular image is masked by the next stimulus in the sequence. Accordingly, low-contrast fine-scale information could be masked by the subsequent presentation of high-contrast coarse-scale information more than vice versa, resulting in a bias for low-to-high sequences.

Schyns and Oliva (1994) used a different technique (hybrid stimuli) to provide evidence of a coarse-to-fine bias in scene processing. Hybrids depict the LSFs from one image and the HSFs from another. This is achieved by superimposing a low-passed image with a high-passed stimulus. Figure 2 shows hybrid stimuli similar to those of Schyns and Oliva (1994). The HSFs represent a city in the left picture and a highway in the right picture. If you squint, blink, defocus, or step away from the pictures, the LSFs information would appear. The LSFs represent the opposite interpretations of a highway in the left picture and a city in the right picture.

For their first experiment, Schyns and Oliva (1994) used a matching task whereby a sample was presented for either 30 or 150 msec, followed immediately by a mask and then a target. Subjects indicated whether or not the sample matched the target. The samples were full-

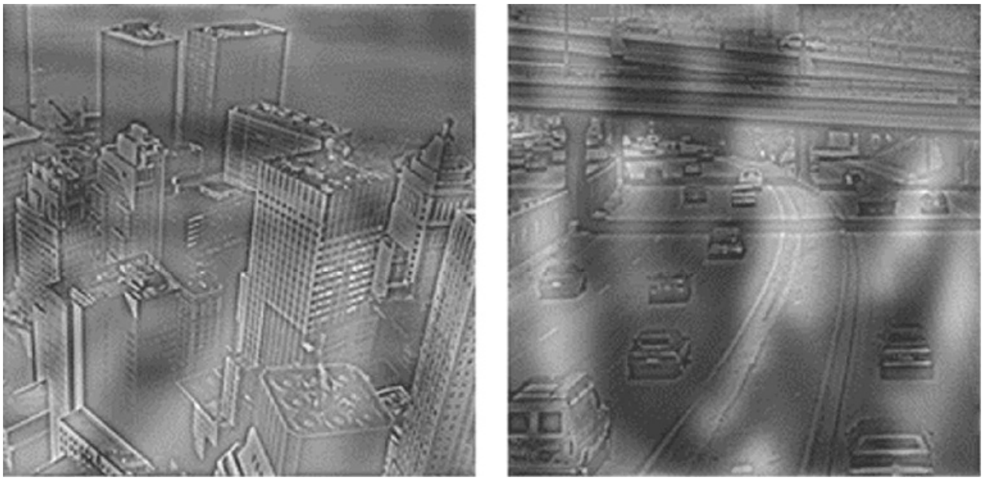

Figure 2. Examples of hybrid stimuli, similar to those used by Schyns and Oliva (1994). The image on the left depicts a city at high spatial frequencies and a motorway at low spatial frequencies (LSFs). Conversely, a city is represented in the coarse blobs of the image on the right, and the boundary edges depict a motorway. To see the LSF content, squint, blink, or step back from the figure.

spectrum, low-passed, high-passed, or hybrid images, and targets were always full-spectrum scenes. For LSF hybrids, the low frequencies matched the target (i.e. the LSFs of the hybrid represent the same scene as the full-spectrum target), and for HSF hybrids, the high frequencies matched the target. Thus, a single hybrid could be matched with two different scenes, one depicted in LSFs and another in HSFs. The two scenes represented by one hybrid could both be matched with their respective target at 30- and 150-msec durations. Nevertheless, exposure duration changed the interpretation of the hybrids: Short exposures elicited more accurate matchings of LSF hybrids, as compared with the long exposures, whereas the converse was true of HSF hybrids. This finding in a scene-matching task is consistent with a coarse-to-fine mode of processing. Matching tasks, however, are very different from typical situations of categorization, and they tap into different processes (e.g., Biederman & Cooper, 1992). In a second experiment Schyns and Oliva (1994) obtained evidence for a coarse-to-fine recognition (as opposed to matching) strategy. Each trial was an animation created by the sequential presentation of two hybrids for 45 msec each without interval. An animation contained two distinct sequences, one coarse-to-fine and the other fine-to-coarse—that is, observers saw two different scene sequences simultaneously. For example, if the left hybrid from Figure 2 is immediately followed by that on the right, the coarse-to-fine sequence would represent a motorway, whereas the fine-to-coarse animation would depict a city. When asked to name the scene in the sequence, observers chose the coarse-to-fine interpretation more frequently than the fine-to-course scenario (67% vs. 29%, respectively). This is evidence in support of a coarse-to-fine categorization strategy. Despite this, later evidence suggests that categorization sometimes proceeds in the opposite direction. Before considering this evidence, we turn to a seemingly related literature concerned with the global-to-local phenomenon.

Coarse-to-Fine and Global-to-Local

There is an apparent analogy between coarse-to-fine processing and a phenomenon called *global-to-local* (e.g., Hughes, 1986; Hughes, Fendrich, & Reuter-Lorenz, 1990; Hughes, Nozawa, & Kitterle, 1996; Lamb & Yund, 1993, 1996a, 1996b; Navon, 1977; Paquet & Merikle, 1988; Robertson, 1996; among many others). To illustrate, Navon used hierarchically organized letters similar to those in Figure 3 (adapted from Oliva & Schyns, 1997). He found that whereas the global processing of F was not affected by the local Ls, the incongruent global letters hindered the local processing of Ls. This asymmetry, called the *global precedence effect*, predicts that global structures in an image are generally processed before local structures; the forest precedes the trees (Navon, 1977).

Several authors have proposed a general link between the global precedence effect and coarse-to-fine processing (Badcock, Whitworth, Badcock, & Lovegrove, 1990; Hughes et al., 1996; Lamb & Yund, 1996b; Shulman, Sullivan, & Sakoda, 1986; Shulman & Wilson, 1987): The temporal delay between LSF and HSF channels discussed earlier could explain the precedence of global information (e.g., Breitmeyer, 1984; Ginsburg, 1986; Marr, 1982; Parker et al., 1992; among many others).

Does coarse-to-fine-scale processing provide the supporting mechanisms and representations of global-to-local, to the extent that the latter is reducible to the former? We believe that this conclusion might be premature. To illustrate, consider Figure 3. This hybrid comprises global letters: HSFs represent F (for fine), and LSFs represent C (for coarse). The fact that you *can* read F and C demonstrates that global processing can occur at both the coarse *and* the fine scales. A closer look at the hybrid reveals that both C and F are composed of smaller Ls (for local). The fact that you *can* read the Ls demonstrates that local processing can also be accomplished at the coarse *and* the fine scales. On this account, coarse-to-fine is a processing mode conceivably orthogonal to global-to-local (Oliva & Schyns, 1997). Global-to-local operates in the two-dimensional (2-D) image plane, depending on the spatial extent of the 2-D image information that is integrated (e.g., most of the visual field or just a small part). Coarse-to-fine occurs in another, *n*-dimensional scale space. To picture the proper relationship between these spaces, imagine a third axis orthogonal to the image plane. This axis represents *n* 2-D image planes (one per scale; in Figure 3, $n = 2$). In the analogy, coarse-to-fine is a process that takes place along the third dimension, whereas global-to-local operates at any of the *n* 2-D planes. This idea, without the mandatory sequencing, is the essence of wavelet analysis (Mallet, 1989, 1991).

At this juncture, we mention work that appears to contradict a systematic strategy of categorizing coarse-scale structures before fine-scale details. In physiology, it appears that the spatiotemporal properties of transient and sustained channels, on which the argument of a temporal delay between LSF and HSF is based, are contentious. As was stated in de Valois and de Valois (1990, p. 111): “There is no evidence, either within the simple cell population or within the complex cells or within the population as a whole, for a bimodal distribution of temporal properties such as would justify a dichotomy into sustained versus transient cell types. Furthermore, a comparison of the temporal properties of simple versus complex cells also indicate little evidence for any significant temporal difference between these two classes of cells, which differ so drastically in their spatial properties.” In other words, an eventual bias from physiology would not be so constraining as to impose a coarse-to-fine recognition scheme.

In recognition, Parker, Lishman, and Hughes (1996) examined how coarse- and fine-scale information guides the processing of complex visual stimuli, using a same–different matching task. To do so, samples on some trials comprised a filtered (low- or high-passed) image (100

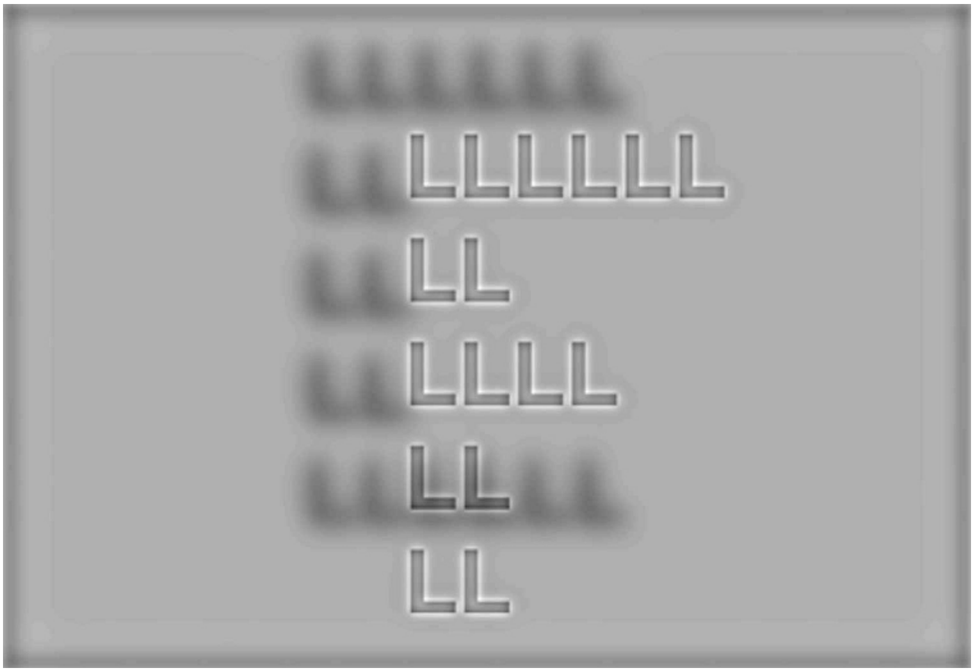

Figure 3. One possible difference between coarse-to-fine and global-to-local. Low spatial frequencies represent the letter C (for coarse) and high spatial frequencies represent the letter F (for fine). The reading of C and F demonstrates that global processing is possible at the coarse and fine scales. The reading of the small Ls (for local) composing the global letters demonstrate that local processing is also possible at coarse and fine scales. Hence, coarse-to-fine and global-to-local could operate orthogonally to one another. Global-to-local takes place in the two-dimensional visual field, whereas coarse-to-fine occurs on a third dimension, orthogonal to the image plane—in the picture, this third dimension comprises two different image planes. From "Coarse Blobs or Fine Edges? Evidence that Information Diagnosticity Changes the Perception of Complex Visual Stimuli," by A. Oliva and P. G. Schyns, 1997, *Cognitive Psychology, 34*, p. 77. Copyright 1997 by Academic Press. Adapted with permission.

msec) followed by a full-spectrum image (400 msec). Using both faces and objects with very different global properties, Parker et al. (1996, p. 1452) concluded that any biases were in favor of high-passed images: "The pattern of results found in all experiments lends no support to the view that the natural path to object recognition is initially via coarse-scale information." However, as was mentioned already, matching and recognition are quite different tasks. Still, the evidence is not in favor of a mandatory coarse-to-fine recognition strategy. Oliva and Schyns (1997, Experiment 1) *did* use an identification task and demonstrated that the LSF and the HSF components of a hybrid scene (presented for 30 msec) both primed subsequent recognition of a full-spectrum scene. Therefore, both coarse- and fine-scale cues are available early, arguing against a mandatory, perceptually driven coarse-to-fine recognition scheme. In the global-to-local literature, several researchers have demonstrated that the global precedence effect was itself not systematic but, instead, modulated by task constraints. For example, Grice, Canham, and Boroughs (1983) showed that an advantage for the global interpretations of larger letters made of smaller letters could be overcome when subjects could attend to and fixate the local constituent letters (see also Kimchi, 1992, and Sergent, 1982, for reviews).

A Flexible Categorization Strategy?

According to the flexible usage hypothesis, categorization mechanisms tune into the scales that convey task-relevant, or diagnostic, information. For example, the age of a face can be assessed from the wrinkles around the eyes and mouth, the sharpness of its contours, and other such local cues that are poorly represented at a coarse scale. One would hypothesize that the age of a face would be better determined from fine-scale information, suggesting that categorization mechanisms should tune preferentially to information present at this scale. In contrast, specific face expressions (e.g., happiness) are more global and quite resistant to changes of scale, suggesting that they are already well represented at a lower spatial scale. One could expect that subjects categorizing this expression would preferentially use information at a coarse scale. This argument of a flexible scale use is not limited to faces; it applies to any visual categorization (Schyns, 1998). Two factors need to be considered: the categorization task, which specifies the demands of visual information from the input, and the

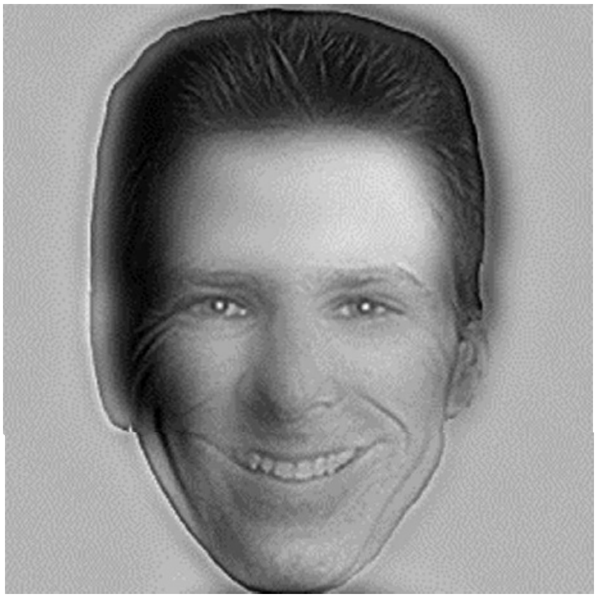

Figure 4. A hybrid used by Schyns and Oliva (1999). The low spatial frequencies depict an angry female and the high spatial frequencies depict a happy male.

representation of this visual information across the scale space.

Data consistent with this flexible stance were reported by Oliva and Schyns (1997) and Schyns and Oliva (1999). Observers in Oliva and Schyns's second experiment were presented with scenes, for 135 msec, for identification (*city*, *highway*, *living room*, or *bedroom*?). They first saw images that were meaningful at LSFs or HSFs only—for example, a fine-scale highway combined with coarse-scale noise. Without discontinuity in presentation, hybrids were presented—for example, HSFs depicted a city, and LSFs a bedroom. Hybrids were identified in accordance with the scale at which diagnostic information was initially presented. That is, those sensitized to fine scales perceived the HSF component from a hybrid, whereas those sensitized to coarse scales perceived the LSF scene from the identical hybrid. The observers claimed to be unaware that two different scenes were depicted in any one image, which argues against the possibility that the observers perceived two scenes in hybrids and decided to report that consistent with the sensitization phase. This finding, which has more recently been replicated with faces of famous people (Morrison & Schyns, 2001), suggests that scale usage is flexible and tunes into the scale at which diagnostic information is conveyed.

Central to the flexible usage hypothesis is the idea that different categorizations of identical visual inputs rely on distinct regions of the spatial spectrum—for example, distinct spatial frequencies may convey face identity, gender, and expression (Sergent, 1986). If this is the case (we return to this topic later), the flexible usage hypothesis predicts that the perception of identical hybrids should depend on the categorization being performed. Schyns and Oliva (1999) addressed this question by using hybrids derived from the faces of unfamiliar people. For example, a happy male at HSFs may be superimposed with an angry female at LSFs (see Figure 4). In their first experiment, stimuli were presented for 50 msec, and the nature of the categorization was indeed found to moderate the perception of the stimulus. For example, when asked whether the face was expressive or not, observers tended to perceive and report the fine-scale face, whereas there was no bias for a gender decision and there was a coarse-scale bias when asked to pinpoint the expression as happy, angry, or neutral. Again, the observers remained unaware of the presence of two faces in any one image. In short, perception of identical hybrids was determined by the categorization task, suggesting that categorization processes tune into diagnostic information at specific scales.

Categorization-dependent scale perceptions are important to understand how the higher level categorization task can modify the lower level parameters of perceptual processes. However, stronger evidence than that just reviewed might be required to corroborate that categorization does indeed modulate scale perception. Schyns and Oliva (1999) designed a second experiment to isolate the perceptual by-products of a categorization task. In a first phase, two subject groups applied a different categorization (expressive or not vs. which expression) to an identical set of hybrid faces, to induce two orthogonal biases (to HSF and LSF, respectively). In the second phase, all the subjects were asked to judge the gender of the same set of hybrid faces. The results established a perceptual transfer of the bias acquired in a first categorization to the subsequent gender task. To illustrate, when one group preferentially categorized the hybrid of Figure 4 as a male on the basis of its HSF, the other group categorized the same picture as a female on the basis of its LSF. Note that the groups differed only on the frequency bandwidth bias acquired in the first phase of the experiment. In the second phase, all aspects of the experimental task (i.e., the gender categorization, the hybrid stimuli, and their conditions of presentation) were strictly identical across subjects, who nevertheless perceived the same hybrid faces quite differently. This perceptual transfer licenses the conclusion that categorization can modify the perception of scale information.

Summary

A commonly held view is that there is a coarse-to-fine bias in the processing of spatial scales (e.g., Marr & Hildreth, 1980; Watt, 1987). It would seem that LSFs are extracted before HSFs from simple stimuli (e.g., Parker & Dutch, 1987) and that scale information may be integrated more efficiently in a coarse-to-fine sequence (Parker et al., 1992, 1997). This itself does not imply the existence of a coarse-to-fine recognition strategy, however. In fact, recent evidence (Morrison & Schyns, 2001; Oliva & Schyns, 1997; Schyns & Oliva, 1999) conflicts with the view that scale usage for categorization is fixed and, rather, suggests that it is flexible and driven by the presence of diagnostic information at different scales. The data that argue for a coarse-to-fine recognition strategy (e.g., Schyns & Oliva, 1994, Experiment 2) do not conflict with the flexible usage hypothesis. According to this position, there are instances in which scale usage may operate in a coarse-to-fine manner. Schyns and Oliva (1994) used only one task (what is this scene?) and may have stumbled across one instance in which coarse-to-fine processing was optimal for this scene recognition task. On the other hand, demonstrations of flexible scale usage do conflict with any strong version of the coarse-to-fine hypothesis. Therefore, although there may be early coarse-to-fine biases for the perceptual extraction and integration of spatial scales, scale usage for categorization is not governed in such a fixed coarse-to-fine manner. Rather, scale usage for categorization appears to be flexible and determined by the diagnosticity of information across the spatial spectrum. Furthermore, there is now converging evidence that the diagnostic use of coarse- and fine-scale cues in categorization tasks does change the perceptual appearance of the incoming stimulus. This could have far-reaching implications for theories of recognition and perception, to which we will return in the General Discussion section when we discuss the specific parameters of spatial scale filtering that could be under cognitive influence.

CATEGORIZATION INFORMATION AND SPATIAL SCALES: THE CASE OF FACE IDENTITY

We have argued in favor of a flexible, rather than fixed, use of spatial scales for categorization tasks. Accordingly, categorization mechanisms tune into the scales that convey task-relevant information. The flexible usage hypothesis is thus based on the assumptions that (1) the perceptual cues important for a particular categorization may be associated with a restricted range of spatial frequencies and (2) different regions of the spatial spectrum are important for distinct categorizations (e.g., gender, identity, expression, and so forth, for a face) of an identical visual input. We now consider the empirical work that bears on these issues.

The question of whether diagnostic cues are associated with a restricted band of the spatial spectrum has been addressed with respect to face recognition. The discovery of a range of spatial frequencies that best transmit identity would have considerable practical implications. For example, recognition algorithms could be purposefully designed in order to learn to identify faces from this most informative and restricted information bandwidth. Compression algorithms (Burt & Adelson, 1983; Lindeberg, 1993; Strang & Nguyen, 1997) could also be built with the knowledge of how to compress images that contain faces—to retain their identity after compression. Given the range of practical tasks in which an algorithm for identifying faces could be used, it is of primary importance to determine the critical information for their identification.

A number of studies, using slightly different techniques, have evaluated which spatial frequencies are particularly important for identifying faces (Bachmann, 1991; Bachmann & Kahusk, 1997; Costen, Parker, & Craw, 1994, 1996; Fiorentini et. al, 1983; Harmon, 1973; Harmon & Julesz, 1973; Parker & Costen, 1999; Parker et al., 1996). Parker and Costen provide a concise summary of this work. The quantization (or pixelating) technique is considered later; here, we focus on studies using low-, high- , and bandpassed images. Everyday face identification is largely an effortless and rapid procedure, since when we encounter the face of a known person under adequate viewing conditions, we have no difficulty recognizing it as being familiar. Of course, retrieving specific stored knowledge or a name via a face is often a procedure that is slower and more susceptible to error (e.g., Young, Hay, & Ellis, 1985). Unfortunately, some of the studies above used procedures, such as face matching, that do not reflect more natural face identification and may tell us very little about the spatial frequencies important for this task (e.g., Harmon, 1973; Hayes, Morrone, & Burr, 1986; Parker et al., 1996). As was already mentioned, this is because different processes underlie matching and recognition (see Biederman & Cooper, 1992; Sergent & Poncet, 1990; Young, Newcombe, de Haan, Small, & Hay, 1993).

Procedures that do appear to tap more natural face recognition mechanisms have also been used. For example, Fiorentini et al. (1983) trained subjects to identify nine originally unfamiliar male faces and examined how recognition was affected when the faces were low- and high-passed. These authors concluded that both coarse- and fine-scale information can be used to identify faces but that a central bandwidth of spatial frequencies is particularly important. However, interpretation of these data is hampered, since response latency was not recorded and assessing the accuracy of face identification without applying time pressure may disguise variations in recognition efficiency (Parker & Costen, 1999). To circumvent this problem, Costen et al. (1994, 1996) recorded both the accuracy and the latency with which subjects could identify low- and high-passed faces and largely agreed with Fiorentini et al. by concluding that a central band of frequencies (about 8–16 cycles per face width, measured at eye level) are particularly important for conveying face identity. Unfortunately, all these studies used stimuli derived from the same image at training and test, so the results may tell us more about picture recognition than about face identification. More recently, however, Parker and Costen trained observers to identify six previously unfamiliar faces from one viewing angle and tested subsequent recognition of these faces when viewed from five different angles. Test stimuli were produced by bandpassing the faces with filters one octave wide. Identification efficiency (again, as indexed by speed and accuracy) was best for the bands centered at 5.22, 11.1, and 23.6, cycles per face width but dropped off when the bands were centered on 2.46 and 50.15 cycles per face width. These results do appear to confirm the view that a central band of spatial frequencies is more useful for identifying faces. Spatial information derived from the face, including the band between 8 and 16 cycles per face width that Parker and Costen conclude is important for face recognition, is shown in Figure 5.

A number of findings indicate that faces can be identified via coarse or fine scales alone, although a central bandwidth appears to be particularly important. Parker and Costen (1999), in particular, avoided some of the pitfalls in this area, since they assessed both accuracy and speed of identification and used different images at training and test. These studies provide support for the first assumption of the flexible usage hypothesis—namely, that the perceptual cues important for a particular categorization (in this case, identifying a small number of recently learned faces) are associated with specific spatial frequencies. Unfortunately, such investigations have only used one categorization (face identification) and so do not address the second assumption of flexible usage. To do so, an experiment should assess how the perception of distinct information from an identical stimulus is influenced by spatial content. For example, under identical viewing conditions, does spatial content influence equivalently the perception of face identity and expression?

Although the results just described do indicate that diagnostic information may be conveyed by a restricted band of spatial information, further claims on the basis of such data must be made with great care. For example, the studies discussed above explored the identification of a small set of recently learned faces. The frequencies that appear to be important for face identification may be influenced by both the size of the target set and the familiarity of these faces. The cues used to distinguish 1 face among 6 might be quite different from those required to distinguish the same face among 60 others. It is also known that familiar and unfamiliar faces are processed in different ways and that the cues used to identify a face may change as the face becomes increasingly

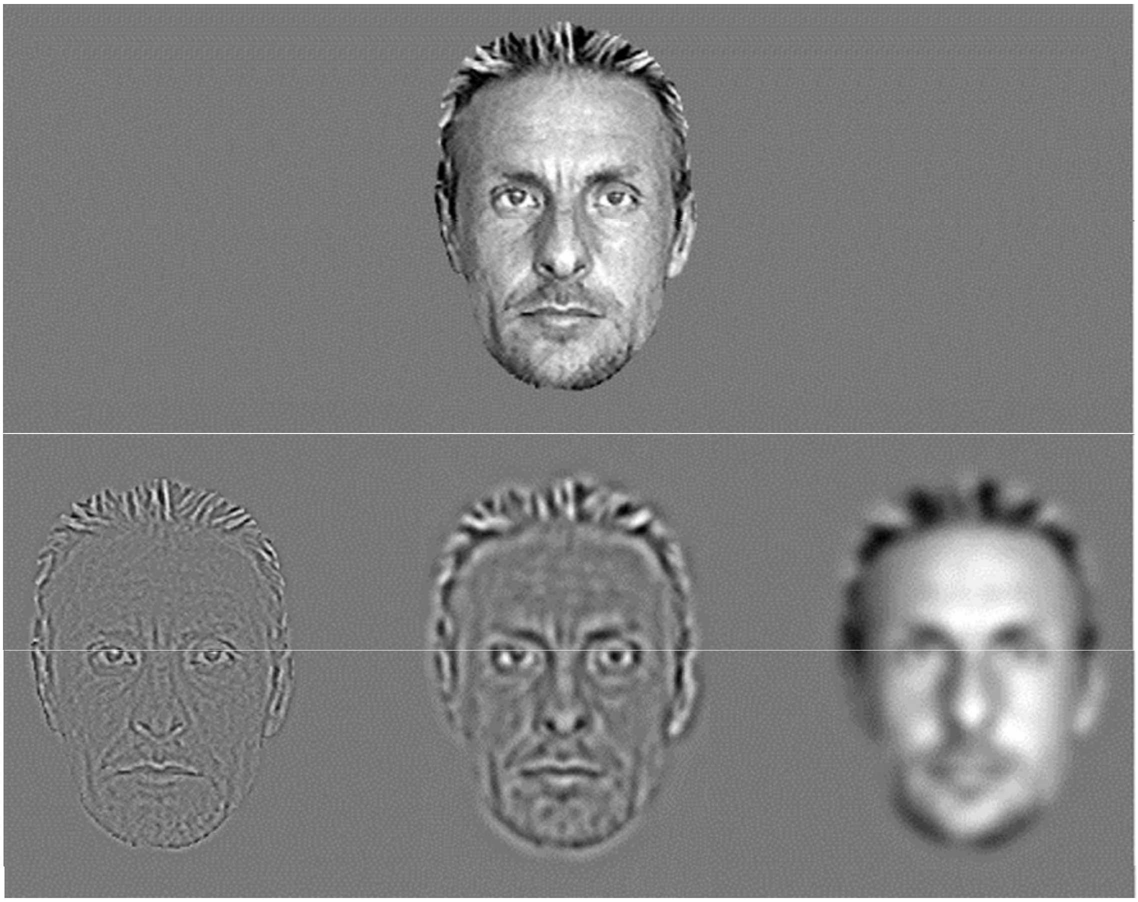

Figure 5. The full-spectrum face (top) was filtered to extract a range of spatial information. Parker and Costen (1999) state that the information between 8 and 16 cycles per face width (bottom middle) is relatively important for conveying face identity. Also shown are those frequencies above 16 cycles per face width (bottom left) and those below 8 cycles per face width (bottom right).

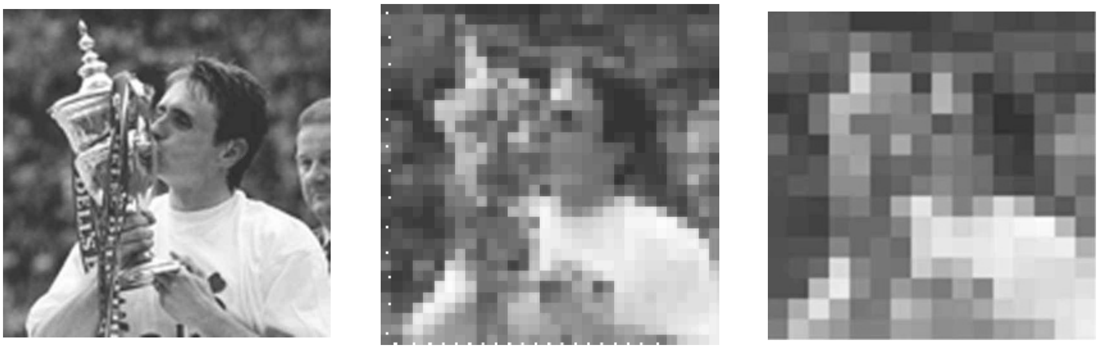

Figure 6. Examples of spatially quantized images derived from the original image (left). The levels of quantization are 32 and 16 pixels per image for the middle and the right items, respectively.

familiar (e.g., O'Donnell & Bruce, 2001). These cues may well be carried by different spatial frequencies, so these factors limit the conclusions that can be drawn from such empirical work.

Quantization

The experiments discussed thus far have employed Fourier filtering techniques to examine the usage of spatial frequencies in the processing of complex visual images. Another method for transforming images, known as spatially quantizing (or blocking or pixelating), has also been used in this area. Spatially quantized images are created by placing a regular square grid across the image and setting the luminance of each grid square to the average luminance within it. Examples of blocked images are depicted in Figure 6. Blocking an image is a low-filtering operation, since higher spatial frequencies are removed (but see below). Since Harmon and Julesz's (1973) classic paper, a number of studies have examined the effect of blocking on the recognition of faces and objects (e.g., Bachmann, 1991; Bachmann & Kahusk, 1997; Costen et al., 1994, 1996; Uttal, Baruch, & Allen, 1995) and Bachmann and Kahusk provide an overview of this work.

Studies using the quantization technique have yielded a number of interesting findings, such as the Harmon and Julesz (1973) phenomenon, whereby recognition of a blocked image improves when it is low-passed (e.g., Bachmann, 1991; Harmon & Julesz, 1973; Uttal et al., 1995). This methodology has also been used to assess the spatial bandwidth important for identifying faces (e.g., Bachmann, 1991; Bachmann & Kahusk, 1997; Costen et al., 1994, 1996). However, this approach raises problems when the usage of spatial frequency information in face and object recognition is considered. Spatial quantization is a filtering technique, since higher spatial frequencies are removed. In this respect, blocking and low-passing are similar filtering operations (compare, e.g., the low-passed and quantized images in Figures 1 and 6). Recognition of both quantized and low-passed images may be impaired since task-relevant fine-scale information has been removed. However blocking, unlike Fourier filtering, introduces further factors that may be detrimental to recognition. First, spatial quantization introduces HSF components into an image (created by the corners and edges of the blocks) that may impair categorization by masking relevant information. Second, quantization produces a compound image comprising both the filtered form of the original picture and the pattern of the mosaic itself. These two components may compete for attentional processes, thus impairing identification. Finally, Bachmann and Kahusk point out that quantization performed at more coarse levels disrupts any configuration in an image as uncertainty about the location of specific features is introduced. These problems associated with the blocking technique can be avoided by using Fourier filtered images. Therefore, we suggest that Fourier rather than blocking filtering operations should be used when the usage of spatial scales in the categorization of natural images is explored. Nevertheless, there are occasions when quantized images can reveal effects that may not be seen with low-passed images. Bachmann and Kahusk demonstrated a counterintuitive effect of precuing attention to the location of quantized images—that is, for coarsely quantized stimuli, precuing location actually impaired performance. This finding may be caused by the fine-scale noise in blocked images and would thus not have been found when low-passed items were used.

GENERAL DISCUSSION

We started this review with the observation that there has been a recent increase in activity in the study of recognition based on information at different spatial scales. As we have explained, distinct visual cues for recognition can reside at different spatial scales, which are themselves processed separately in early vision (by frequency-specific channels). We saw that the use of this information for categorization tasks was not determined by early biases but could, instead, be flexibly adjusted to the requirements of the task at hand. Furthermore, in these circumstances, the perception of the stimulus could depend on the scale information selectively attended.

Implications for Attention and Perception

The empirical work reviewed demonstrates that attention can exert a selective control on the scale information used for categorization (e.g., Oliva & Schyns, 1997). Further evidence of selective and task-dependent processing of visual information can be found in psychophysics. For example, detection of sinusoidal gratings is worse when the spatial frequency varies from trial to trial, as compared with when the same gratings are presented in blocks of constant spatial frequency. Such uncertainty effects are consistent with the notion of selective activation or monitoring of spatial frequency channels (Hübner, 1996).

The common underpinnings between the hybrid methodology and the spatial filtering techniques ubiquitous in the psychophysics of early vision provide one promising research avenue for unraveling the precise influence that the categorization task can exert on the perception of a face, object, or scene. For example, one could design a study combining hybrid categorization with psychophysical techniques for understanding whether attention to a diagnostic spatial scale (or neglect of a scale) affects the filtering properties of the earliest stages of visual processing—for example, contrast thresholds, frequency tuning, orientation selectivity.

Recent studies of P. Snowden and Schyns (2000) have started to examine the visual implementation of selective, scale-specific extraction of visual cues. In a within-subjects design, observers were trained to detect near-threshold contrasts in LSF and HSF gratings—low and high gratings were cued with a distinct tone. They found a decrement in grating detection when observers were miscued (e.g., when the LSF tone was followed by an HSF grating), supporting the occurrence of an expectancy effect. Schyns and Oliva (1999) argued that the categorization task could likewise cue people to scale-specific face, object, and scene features. The cuing in Sowden and Schyns suggests one possible implementation of the categorization-dependent perceptions reported in hybrids: Contrast modulation could occur in spatial frequency channels as a function of task-related expectations, enhancing or lowering the availability of scale-specific information for subsequent processing. Evidence that categorization tasks can exert such influence would have far-reaching implications for classical issues in cognitive science, ranging from the depth of feedback loops in early vision, the early versus late selection models of attention (Pashler, 1998), the bidirectionality of cognition (Schyns, 1998), the sparse versus exhaustive perceptions of distal stimuli (Hochberg, 1982), to the cognitive penetrability of vision (Fodor, 1983; Pylyshyn, 1999).

A striking observation in studies with hybrid stimuli is that people who are induced to attend and, consequently, perceive consciously information depicted at only one scale appear to be unaware of some aspects of the cues at the other scale. This leads to the question of whether unattended scale information is nevertheless recognized covertly and, if so, at what level of specificity? For example, in a recent study (Morrison & Schyns, 2001), two groups of observers were initially sensitized to identify the faces of famous people at either LSFs or HSFs (the other scale was noise). After a few trials and without the subjects being told of a change, hybrids were presented that depicted the faces of two different celebrities, one at fine and the other at coarse scales. Both LSF and HSF groups performed similarly with respect to identifying the faces in hybrids: The observers recognized the face at the sensitized scale accurately and claimed to be unaware of the identity of the face at the unattended scale. However, the groups differed, since the observers sensitized to HSFs detected the face at the unattended scale (for them, the coarse scale face) more accurately than did those in the group sensitized to LSFs (in their case, the fine-scale face). This suggests that people can only perform a precise overt identification at the scale they attend, although cues at the other scale may permit other categorizations, such as detection, and it is possible they are also identified covertly. Similar issues have been addressed in attention research (see Pashler, 1998). The added twist here is that different categorization tasks can be accomplished selectively with attended and unattended information.

Relational and Part-Based Encoding

Visual information gleaned from the world around us varies on a number of dimensions. One such dimension is spatial frequency, and as we have explained, coarse blobs and fine edges are very different sorts of recognition information. Another dimension refers to whether cues are encoded in a part-based (piecemeal) or a relational (wholistic) manner. Furthermore, these modes of processing may be associated with different spatial scales, as we highlight by focusing on work in the face-processing literature.

It is widely accepted that face processing may rely on both componential cues (i.e., local features, such as the mouth, the nose, the eyes, or a mole) and noncomponential information (the spatial relations between these features), although how these cues are integrated remains unclear (e.g., Bartlett & Searcy, 1993; Calder, Young, Keane, & Dean, 2000; Farah, Wilson, Drain, & Tanaka, 1998; Macho & Leder, 1998). We use the term *relational* to refer to a mode of processing that encodes the spatial relations of the face without making further claims about the nature of this encoding. Relational and component cues are different sorts of information, since, for example, turning a face upside down has a greater detrimental effect on encoding of the former (e.g., Bartlett & Searcy, 1993; Leder & Bruce, 1998). They may thus be associated with different spatial scales. Indeed, Sergent (1986, pp. 23–24) has argued that "a face has both component and configurational properties that coexist, the latter emerging from the interrelationships among the former. These properties are not contained in the same spatial-frequency spectrum. . . ." More precisely, Sergent

(1986) suggested that component and relational properties may be associated with fine and coarse scales, respectively.

Surprisingly, there are no published studies exploring different modes of face encoding across the spatial spectrum. The suggestion that relational and componential cues may be associated with coarse and fine scales, respectively, could be examined in a number of ways. Consider a finely balanced hybrid depicting one face at LSFs and another face at HSFs. When presented upright for categorization, relational encoding should be implicated, so we may expect a bias in favor of LSFs. On the other hand, when the same image is inverted, encoding should be more feature based, resulting in a bias toward perceiving the HSF face. That is, simply rotating a hybrid through 180° in the picture plane should influence whether the coarse- or the fine-scale component will be perceived. Spatial filtering techniques could also be combined nicely with some of the methods used to demonstrate the relational processing of faces, such as the composite effect (e.g., Calder et al., 2000; Young, Hellawell, & Hay, 1987), the face inversion effect (see Valentine, 1988), and the Margaret Thatcher illusion (Bartlett & Searcy, 1993; Thompson, 1980). In fact, recent work (Morrison & Schyns, 2000) has demonstrated that the Margaret Thatcher illusion is stronger for low- than for high-passed faces, providing some support for Sergent's (1986) view.

Tasks, Spatial Content, and Size

There is an important relationship between spatial content and size. Images of different size may vary not only on the basis of specific metrics, but also in terms of spatial content. This is because fine contours (fine-scale information) are better represented in large images, as compared with smaller versions (which comprise only the coarse-scale information of the larger image). For example, to use faces again, certain judgments of expressions (e.g., happiness) are more resilient to changes in viewing distance than are others (see Jenkins, Craven, Bruce, & Akamatsu, 1997). More generally, it will be interesting to examine how different categorization tasks of the same face, such as its gender, expression, age, identity, and so forth, specifically degrade with progressive increases in viewing distance. This will provide a better indication of the scale at which the information necessary to perform this categorization resides (particularly so if the degradation of performance is not linear with the decrease in stimulus size).

A similar reasoning applies to common object and scene categorizations. It is well known that people can apply categorizations at different levels of abstraction to the same stimulus (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976; for a review, see Murphy & Lassaline, 1997). For example, the same animal can be called *collie* at the subordinate level, *dog* at the basic level, and *animal* at the superordinate level. Of these three main levels, two (the basic and subordinate) are arguably closer to perception (see Schyns, 1998, for arguments). The categorization literature has often reported that people seem to be biased to the basic level (Jolicœur, Gluck, & Kosslyn, 1984; Murphy, 1991; Murphy & Smith, 1982; Rosch et al., 1976; Tanaka & Taylor, 1991). The nature of this bias remains a controversy. One possibility is to consider that in natural viewing conditions, we experience objects at many different distances. If, for example, basic-level categorizations were more resilient to changes of scale and viewing distances than were subordinate categorizations, the cues subtending the basic level would be present in most retinal projections of distal objects. This natural bias in the distribution of image cues could bias categorization processes to the basic level, suggesting an interaction between categorization tasks and the differential availability of their scale information.

Archambault, Gosselin, and Schyns (2000) confirmed this hypothesis. In a first experiment, subjects were asked whether two simultaneously presented objects (computer-synthesized three-dimensional [3-D] animals from eight different species, *bird*, *cow*, *dog*, *horse*, *frog*, *turtle*, *spider,* and *whale*, rendered in 256 gray-levels with a Gouraud shading model) had the same basic-level (e.g., whale) or the same subordinate-level (e.g., Humpback whale) category. Object pairs could appear in any one of five sizes, corresponding to 12°, 6°, 3°, 1.5°, 0.75°, and 0.38° of visual angle on the screen. Note that the subjects could inspect the object pairs for as long as they wished, licensing the conclusion that the task was tapping into the absolute level of scale information required for the categorizations. In these conditions, the authors found that subordinate judgments were significantly more impaired by a reduction in stimulus size than were basic judgments. Their second experiment confirmed the results in a straightforward naming task. Thus, constraints on the 2-D proximal projection of 3-D distal objects differentially modify the availability of scale-specific information for basic and subordinate categorizations.

In the flexible usage scenario, the requirements of information arising from different categorization tasks determine a bias to the scale at which these cues are best represented. The experiments just reviewed suggest a natural bias for the finer scales in subordinate categorizations, whereas all scales are equally usable for basic categorizations. This suggests that basic categories are represented in memory either with shape cues that intersect all scales (e.g., a silhouette) or with different cues specific to each scale. In general, we believe that the interactions between the task demands of different categorizations and the structure of input information can selectively modulate the relative needs of visual information at different spatial scales (coarse vs. fine) and spatial extents (global vs. local).

CONCLUDING REMARKS

Different spatial scales can be used for different categorizations of the same face, object, or scene. From our review of the literature, a view emerges that the mechanisms of categorization can modulate the usage of different spatial scales, according to the presence of task-dependent, diagnostic information. Further research is required to unravel the nature of this diagnostic information, for different categorization tasks and the same object, and how this information depends on scale (see Gosselin & Schyns, 2001, for a technique revealing a task-dependent use of scale information). The interactions between these factors could shed a new light on face, object, and scene perception and representation.

REFERENCES

Archambault, A., Gosselin, F., & Schyns, P. G. (2000). A natural bias for the basic level? In *Proceedings of the XXII Meeting of the Cognitive Science Society* (pp. 60-65). Hillsdale, NJ: Erlbaum.

Bachmann, T. (1991). Identification of spatially quantised tachistoscopic images of faces: How many pixels does it take to carry identity? *European Journal of Cognitive Psychology*, **3**, 85-103.

Bachmann, T., & Kahusk, N. (1997). The effect of coarseness of quantisation, exposure duration, and selective spatial attention on the perception of spatially quantised ("blocked") visual images. *Perception*, **26**, 1181-1196.

Badcock, J. C., Whitworth, F. A., Badcock, D. R., & Lovegrove, W. J. (1990). Low-frequency filtering and the processing of local–global stimuli. *Perception*, **19**, 617-629.

Bartlett J. C., & Searcy, J. H. (1993). Inversion and configuration of faces. *Cognitive Psychology*, **25**, 281-316.

Biederman, I., & Cooper, E. E. (1992). Size invariance in visual object priming. *Journal of Experimental Psychology, Human Perception & Performance*, **18**, 121-133.

Blackemore, C., & Campbell, F. W. (1969). On the existence of neurons in the human visual system selectively sensitive to the orientation and size of retinal images. *Journal of Physiology*, **203**, 237-260.

Breitmeyer, B. G. (1975). Simple reaction time as a measure of the temporal response properties of transient and sustained channels. *Vision Research*, **15**, 1411-1412.

Breitmeyer, B. G. (1984). *Visual masking: An integrative approach.* New York: Oxford University Press.

Burt, P., & Adelson, E. H. (1983). The Laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, **31**, 532-540.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception & Performance*, **26**, 527-551.

Campbell, F. W., & Green, D. G. (1965). Optical and retinal factors affecting visual resolution. *Journal of Physiology*, **181**, 576-593.

Campbell, F. W., & Robson, J. G. (1968). Application of the Fourier analysis to the visibility of gratings. *Journal of Physiology*, **197**, 551-556.

Costen, N. P., Parker, D. M., & Craw, I. (1994). Spatial content and spatial quantisation effects in face recognition. *Perception*, **23**, 129-146.

Costen, N. P., Parker, D. M., & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics*, **58**, 602-612.

Davidson, M. L. (1968). Perturbation approach to spatial brightness interaction in human vision. *Journal of the Optical Society of America*, **58**, 1300-1309.

DeValois, R. L., & DeValois, K. K. (1990). *Spatial vision.* New York: Oxford University Press.

Enroth-Cugell, C., & Robson, J. D. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology*, **187**, 517-522.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. W. (1998). What is "special" about face perception? *Psychological Review*, **105**, 482-498.

Findlay, J. M., Brogan, D., & Wenban-Smith, M. G. (1993). The spatial signal for saccadic eye movements emphasizes visual boundaries. *Perception & Psychophysics*, **53**, 633-641.

Fiorentini, A., Maffei, L., & Sandini, G. (1983). The role of high spatial frequencies in face perception. *Perception*, **12**, 195-201.

Fodor, J. (1983). *The modularity of mind.* Cambridge, MA: MIT Press.

Ginsburg, A. P. (1986). Spatial filtering and visual form perception. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.). *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 1-41). New York: Wiley.

Gish, K., Shulman, G. L., Sheehy, J. B., & Leibowitz, H. W. (1986). Reaction times to different spatial frequencies as a function of detectability. *Vision Research*, **26**, 745-747.

Gosselin, F., & Schyns, P. G. (2001). Bubbles: A new technique to reveal the use of visual information in recognition tasks. *Vision Research*, **41**, 2261-2271.

Graham, N. (1980). Spatial frequency channels in human vision: Detecting edges without edges detectors. In C. S. Harris (Ed.), *Visual coding and adaptability.* Hilldsale, NJ: Erlbaum.

Grice, G. R., Canham, L., & Boroughs, J. M. (1983). Forest before trees? It depends where you look. *Perception & Psychophysics*, **33**, 121-128.

Harmon, L. D. (1973). The recognition of faces. *Scientific American*, **229**, 71-82.

Harmon, L. D., & Julesz, B. (1973). Masking in visual recognition: Effects of two-dimensional filtered noise. *Science*, **180**, 1194-1197.

Hayes, T., Morrone, M., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception*, **15**, 595-602.

Henning, G. B., Hertz, B. G., & Broadbent, D. E. (1975). Some experiments bearing on the hypothesis that the visual system analyzes spatial patterns in independent bands of spatial frequency. *Vision Research*, **15**, 887-897.

Hochberg, J. (1982). How big is a stimulus? In J. Beck (Ed.), *Organization and representation in perception* (pp. 191-217). Hillsdale, NJ: Erlbaum.

Hubel, D. H., & Wiesel, T. N. (1977). Functional architecture of macaque visual cortex. *Proceedings of the Royal Society of London: Series B*, **198**, 1-59.

Hübner, R. (1996). Specific effects of spatial-frequency uncertainty and different cue types on contrast detection: Data and models. *Vision Research*, **36,** 3429-3439.

Hughes, H. C. (1986). Asymmetric interference between components of suprathreshold compound gratings. *Perception & Psychophysics*, **40**, 241-250.

Hughes, H. C., Fendrich, R., & Reuter-Lorenz, P. A. (1990). Global versus local processing in the absence of low spatial frequencies. *Journal of Cognitive Neurosciences*, **2**, 272-282.

Hughes, H. C., Nozawa, G., & Kitterle, F. (1996). Global precedence, spatial frequency channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, **8**, 197-230.

Jenkins, J., Craven, B., Bruce, V., & Akamatsu, S. (1997). *Methods for detecting social signals from the face* (Tech. Rep. of IECE, HIP96-39). Kyoto, Japan: The Institute of Electronics, Information and Communication Engineers.

Jolicœur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connection. *Cognitive Psychology*, **19**, 31-53.

Kimchi, R. (1992). Primacy of wholistic processing and global/local paradigm: A critical review. *Psychological Bulletin*, **112**, 24-38.

Lamb, M. R., & Yund, E. W. (1993). The role of spatial frequency in the processing of hierarchically organized stimuli. *Perception & Psychophysics*, **54**, 773-784.

Lamb, M. R., & Yund, E. W. (1996a). Spatial frequency and attention: Effects of level-, target-, and location-repetition on the processing of global and local forms. *Perception & Psychophysics*, **58**, 363-373.

Lamb, M. R., & Yund, E. W. (1996b). Spatial frequency and the interference between global and local levels of structure. *Visual Cognition*, **3**, 401-427.

Leder, H., & Bruce V. (1998). Local and relational effects of distinctiveness. *Quarterly Journal of Experimental Psychology*, **51A**, 449-473**.**

Legge, G. E., & Gu, Y. (1989). Stereopsis and contrast. *Vision Research*, **29**, 989-1004.

Lindeberg, T. (1993). Detecting salient blob-like images structures and their spatial scales with a scale-space primal sketch: A method for focus-of-attention. *International Journal of Computer Vision*, **11**, 283-318.

Macho S., & Leder, H. (1998). Your eyes only? A test of interactive influence in the processing of facial features. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1486-1500.

Mallet, S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Pattern Analysis and Machine Intelligence*, **11**, 674-693.

Mallet, S. G. (1991). Zero-crossings of a wavelet transform. *IEEE Information Theory*, **37**, 1019-1033.

Marr, D., (1982). *Vision.* San Francisco: Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London: Series B*, **207**, 187-217.

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

Marshall, J. A., Burbeck, C. A., Ariely, J. P., Rolland, J. P., & Martin, K. E. (1996). *Journal of the Optical Society of America A*, **13**, 681-688.

Mihaylova, M., Stomonyakov, V., & Vassilev, A. (1999). Peripheral and central delay in processing high spatial frequencies: Reaction time and VEP latency studies. *Vision Research*, **39**, 699-705.

Morgan, M. J. (1992). Spatial filtering precedes motion detection. *Nature*, **355**, 344-346.

Morrison, D. J., & Schyns, P. G. (2000). *Spatial scales in the Margaret Thatcher illusion.* Manuscript submitted for publication.

Morrison, D. J., & Schyns, P. G. (2001). *Less than meets the eye: Interactions between face processing, selective attention and scale perception.* Manuscript submitted for publication.

Murphy, G. L. (1991). Parts in object concepts: Experiments with artificial categories. *Memory & Cognition*, **19**, 423-438.

Murphy, G. L., & Lassaline, M. E. (1997). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts & D. R. Shanks (Eds.), *Knowledge, concepts and categories: Studies in cognition* (pp. 93-131). Cambridge, MA: MIT Press.

Murphy, G. L., & Smith, E. E. (1982). Basic level superiority in picture categorization. *Journal of Verbal Learning & Verbal Behavior*, **21**, 1-20.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, **9**, 353-383.

O'Donnell, C., & Bruce, V. (2001). Familiarisation with faces selectively enhances sensitivity to changes made to the eyes. *Perception*, **30**, 755-764.

Oliva, A., & Schyns, P. G. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology*, **34**, 72-107.

Pantle, A., & Sekuler, R. (1968). Size detecting mechanisms in human vision. *Science*, **162**, 1146-1148.

Paquet, L., & Merikle, P. M. (1988). Global precedence in attended and nonattended objects. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 89-100.

Parker, D. M., & Costen, N. P. (1999). One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. *Current Psychology*, **18**, 118-127.

Parker, D. M., & Dutch, S. (1987). Perceptual latency and spatial frequency. *Vision Research*, **27**, 1279-1283.

Parker, D. M., Lishman, J. R., & Hughes, J. (1992). Temporal integration of spatially filtered visual images. *Perception*, **21**, 147-160.

Parker, D. M., Lishman, J. R., & Hughes, J. (1996). Role of coarse and fine spatial information in face and object processing. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 1448-1466.

Parker, D. M., Lishman, J. R., & Hughes, J. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, **26**, 1169-1180.

Parker, D. M., & Salzen, E. A. (1977). Latency changes in the human visual evoked response to sinusoidal gratings. *Vision Research*, **17**, 1201-1204.

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Pylyshyn, Z. (1999). Is vision continuous with cognition? The case for cognitive impenetrability of visual perception. *Behavioral & Brain Sciences*, **22**, 341-423.

Robertson, L. C. (1996). Attentional persistence for features of hierarchical patterns. *Journal of Experimental Psychology: General*, **125**, 227-249.

Rosch, E., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, **8**, 382-439.

Schyns, P. G. (1998). Diagnostic recognition: Task constraints, object information and their interactions. *Cognition*, **67**, 147-179.

Schyns, P. G., Goldstone, R. L., & Thibaut, J. P. (1998). The development of features in object concepts. *Behavioral & Brain Sciences*, **21**, 17-41.

Schyns, P. G., & Oliva, A. (1994). From blobs to boundary edges: Evidence for time- and spatial-scale-dependent scene recognition. *Psychological Science*, **5**, 195-200.

Schyns, P. G., & Oliva, A. (1999). Dr. Angry and Mr. Smile: When categorization flexibly modifies the perception of faces in rapid visual presentations. *Cognition*, **69**, 243-265.

Sergent, J. (1982). Theoretical and methodological consequences of variations in exposure duration in visual laterality studies. *Perception & Psychophysics*, **31**, 451-461.

Sergent, J. (1986). Microgenesis of face perception. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. M. Young (Eds.), *Aspects of face processing*. Dordrecht: Martinus Nijhoff.

Sergent, J. & Poncet, M. (1990). From covert to overt recognition of faces in a prosopagnosic patient. *Brain*, **113**, 989-1004.

Shulman, G. L., Sullivan, M. A., Gish, K., & Sakoda, W. J. (1986). The role of spatial-frequency channels in the perception of local and global structure. *Perception*, **15**, 259-273.

Shulman, G. L., & Wilson, J. (1987). Spatial frequency and selective attention to local and global information. *Perception*, **16**, 89-101.

Snowden, R. J., & Hammett, S. T. (1992). Subtractive and divisive adaptation in the human visual system. *Nature*, **355**, 248-250.

Snowden, P., & Schyns, P. G. (2000). Expectancy effects on spatial frequency processing: A psychophysical analogy to task-dependent processing of "real-world" objects and scenes. *Perception*, **29**, 24.

Strang, G., & Nguyen, T. (1997). *Wavelets and filter banks.* Wellesley, MA: Wellesley-Cambridge Press.

Tanaka, J. W., & Taylor, M. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology*, **23**, 457-482.

Thomas, J. P. (1970). Model of the function of receptive fields in human vision. *Psychological Review*, **77**, 121-134.

Thompson, P. (1980). Margaret Thatcher: A new illusion? *Perception*, **9**, 483-484.

Uttal, W. R., Baruch, T., & Allen, L. (1995). Combining image degradations in a recognition task. *Perception & Psychophysics*, **57**, 682-691.

Valentine, T. (1988). Upside-down faces: A review of the effects of inversion upon face recognition. *British Journal of Psychology*, **79**, 471-491.

Watt, R. J. (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 A*, **4**, 2006-2021.

Webster, M. A., & De Valois, R. L. (1985). Relationship between spatial frequencies and orientation tuning of striate-cortex cells. *Journal of the Optical Society of America A*, **2**, 1124-1132.

Wilson, H. R., & Bergen, J. R. (1979). A four-mechanism model for spatial vision. *Vision Research*, **19**, 1177-1190.

Witkin, A. P. (1987). Scale-space filtering. In M. A. Fischler & O. Firschein (Eds.), *Readings in computer vision: Issues, problems, principles and paradigms* (pp. 329-332). Los Altos, CA: Morgan Kaufmann.

Young, A. W., Hay, D. C., & Ellis, A. W. (1985). The faces that launched a thousand slips: Everyday difficulties and errors in recognising people. *British Journal of Psychology*, **76**, 495-523.

Young, A. W., Hellawell, D. J., & Hay, D. C. (1987). Configural information in face perception. *Perception*, **16**, 747-759.

Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury: Selective impairments affecting identity and expression. *Brain*, **116**, 941-959.

NOTES

1. From the outset, it is worth pointing out that we will here use categorization and recognition interchangeably. In our view, 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 (see Schyns, 1998, for further discussions).

2. To illustrate the notion of a cycle, imagine an image of 32 × 32 pixels. The highest spatial frequency it can represent is 16 cycles per image, where each cycle comprises a white pixel followed by a black pixel (or vice versa). In the image, the 16 cycles would represent a left-to-right fine-grained zebra crossing—in fact, the finest crossing that the 32 × 32 pixel image can possibly represent. The lowest complete frequency it can represent is 1 cycle per image—the first 16 adjacent pixels represent a grating going from mid-gray to black and back to mid-gray, and the remaining pixels represent a grating going from mid-gray to white and back to mid-gray. The 32 × 32 pixel image can therefore represent frequencies between 1 and 16 cycles per image. The amplitude of each spatial frequency can be modulated. For example, the extrema of the 16 cycles per image (and the 1 cycle per image) could be gray values, instead of black and white.

Real-world images are generalizations of these simple images: They typically comprise spatial frequencies at many different orientations (not just horizontal, as in the zebra crossing example, but also vertical and all diagonals). The Fourier transform specifies exactly how each spatial frequency individually contributes to the complete image (with amplitude coefficients) and how the different spatial frequencies are coordinated to represent the scene (with phase information). A low-passed (vs. high-passed) image only comprises spatial frequencies below (vs. above) a given number of cycles per image. This frequency is called the *cutoff frequency*, the point above (vs. below) which spatial frequencies in all directions of the image (horizontal, vertical, and all diagonals) are filtered out. Technically, their amplitude is set to zero, and so these frequencies have no expression in the filtered image.

(Manuscript received December 29, 1999;
revision accepted for publication January 9, 2001.)