Footnotes: 1
Tables: 0
Figures: 5

Diagnostic Use of Scale Information for Componential and Holistic Recognition

Philippe G. Schyns and Frédéric Gosselin

Running head: Scale-Dependent Componential and Holistic Recognition

Philippe G. Schyns
Dept of Psychology
University of Glasgow
56, Hillhead Street
Glasgow G12 8QB
UK
philippe@psy.gla.ac.uk
phone: 44 141 330 49 37

Frédéric Gosselin
Département de psychologie
Université de Montréal
C.P. 6128 succ. Centre-ville
Montréal QC
H3C 3J7 Canada
[PHIL: mon addresse email etait incorrecte]Frederic.Gosselin@UMontreal.ca
Phone : (514) 343-7550

People can often place an identical visual stimulus into a number of different categories. For example, the top picture of Figure 1 is a woman, with a neutral expression, called "Mary," if this were her identity. In contrast, the bottom picture is a male, with an angry expression, called "John". In a similar vein, the top scene in Figure 2 is an outdoor scene, a city, or New York at increasingly specific levels of categorization. The bottom picture is an outdoor scene, a highway, and only specialists would know that it is I-95. These different judgments of similar images reveal the impressive versatility of categorization, but also its considerable resilience to failure. For example, if you did not know the identity of the faces (or scenes), you could nevertheless categorize them as male or female (or city or highway). Categorization is this fundamental process that progressively reduces highly variable perceptual inputs into a small number of classes of equivalence (called 'categories"—e.g., *face*, *neutral expression*, *female*, *outdoor scene*, *city*, *New York*) whose memory representations mediate thinking and adaptive action.

INSERT FIGURES 1 AND 2 ABOUT HERE

One fundamental problem for recognition theorists is to understand which visual information is used to access categories in memory. Here, we will examine how variations of [PHIL: contrast plutôt que luminance?] (the grey-levels of an image) are used at different scales to recognize and perceive complex visual events such as faces and scenes. Research in psychophysics and neurophysiology indicates that vision breaks down incoming stimuli into a number of spatial scales (or spatial frequencies). Spatial filtering is usually construed as an early stage of visual processing, the outputs of which form a basis for the higher-level operations of face, object and scene recognition. A complete account of recognition will therefore require a good understanding of spatial filtering and the constraints they impose.

Spatial filters encode [PHIL: same here] luminance variability in the visual field. For example, spatial filters operating at a fine spatial resolution (i.e. high spatial frequencies) encode the detailed edges portraying the contours of a nose, eyelashes, the precise shape of the mouth and eyes, and so forth. In contrast, coarser spatial filters (i.e. low spatial frequencies) could encode pigmentation and shape from shading from the face. That is, spatial filters encode a wide range of useful information. How we use information at these scales might therefore have implications for how we categorize everyday face, object and scene categorization from the outputs of spatial filters.

Research on spatial filtering is an established tradition of psychophysics. We will see that the study of scale usage is an excellent medium to examine the interactions between perception and cognition. To illustrate, if the visual cues used for different categorizations of an identical input (face, object or scene) reside at different spatial frequencies, the low-level processing of spatial frequencies could constrain categorization. On the other hand, the categorization task could itself modify 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).

The chapter is organized as follows. To begin, we introduce the key concepts of spatial scales and spatial frequency channels. Theories of scale usage are then reviewed in light of recent empirical findings. New methods that can reveal the features underlying different categorizations are finally described.

SPATIAL SCALES

Natural images provide the viewer with a wide spectrum of spatial information, ranging from coarse to very fine. Fine spatial information tends to be associated with image details, whereas coarse spatial information corresponds to larger, less detailed aspects. We can describe this spectrum of spatial information with Fourier analysis (Campbell & Green, 1965; Davidson, 1968). The coarse spatial information in the image becomes the Low Spatial Frequencies (LSFs) and the fine spatial information the High Spatial Frequencies (HSFs). Examples of low and high spatial frequencies derived from natural images are shown in figures 1 and 2. HSFs reveal a neutral woman and an angry man in Figure 1a and 1b, respectively, and a city and highway in Figure 2a and 2b, respectively. HSFs preserve fine details such as the eyelashes of the faces, the details of their wrinkles, or the windows of the city buildings.

If you squint, blink, defocus or step back from figures 1 and 2, their interpretation turn into an angry man in Figure 1a, and a neutral woman in Figure 1b; a highway in Figure 2a, and a city in Figure 2b. LSFs can be seen to correspond to the coarse, less detailed parts of the pictures: Properties such as the colour and luminance of blobs are carried in the LSFs whereas fine details, such as the eyelashes, are discarded. Luminance blobs provide a skeleton of information from which fine details can be fleshed out.

A spatial frequency channel is a filter that outputs a restricted range of the information it receives in input. Three channel types are often distinguished. A low-pass channel passes all frequencies below a given cut-off while discarding the frequencies above this cut-off. Conversely, a high-pass channel retains the frequencies above a cut-off while discarding those below it. Finally, a band-pass channel only passes the frequencies between two cut-offs, discarding those at each end.

Psychophysical studies have demonstrated that early vision filters natural stimuli into a number of separate channels, each tuned to a specific bandwidth of spatial frequencies (see DeValois & DeValois, 1990, for an excellent review of spatial vision). In their seminal paper on contrast detection, Campbell and Robson (1968) reported that the detection (and the discrimination) of simple sinewave patterns was predicted by the contrast of their frequency components. This could only occur if early vision was analyzing the patterns with groups of quasi-linear band-pass filters, each tuned to a specific frequency band (see also Graham, 1980; Pantle & Sekuler, 1968; Thomas, 1970; Webster & DeValois, 1985). Frequency-specific adaptation studies demonstrated that the channels were selectively impaired in their sensitivity to contrast, suggesting they are independent (e.g. Blakemore & Campbell, 1969).

The visual input appears to be independently processed by 4 to 6 spatial frequency channels (Ginsburg, 1986; Wilson & Bergen, 1979). Further developments indicated that the channels interacted (e.g., Henning, Hertz & Broadbent, 1975) and were nonlinear (e.g., Snowden & Hammett, 1992). In spite of this, the consensual view is that spatial filtering occurs prior to many other early visual tasks such as motion perception (e.g., Morgan, 1992), stereopsis (Legge & Gu, 1989), edge detection (e.g., Marr, 1982; Watt & Morgan, 1985) and saccade programming (Findlay, Brogan & Wenban-Smith, 1993). Spatial filters therefore provide a plausible candidate for the building blocks of visual perception from which flexible categorizations of faces, objects and scenes might arise.

SCALE USAGE FOR CATEGORIZATION

If vision filters the input at multiple spatial scales, the question arises as to how information from these channels is used to categorize complex stimuli. Two scenarios of scale usage are possible. Early constraints on the availability and extraction of coarse and fine scale information may impose a fixed order on their usage in 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 demands of the categorization task at hand.

Fixed usage: Coarse-to-fine hypothesis

A commonly held view is that there is a fixed bias to process scale information from coarse to fine, both in early vision, and in its usage for face, object and scene recognition (e.g., Breitmeyer, 1984; Fiorentini, Maffei & Sandini, 1983; Parker & Costen, 1999; Parker, Lishman & Hughes, 1992, 1997; Schyns & Oliva, 1994). This idea originates in a classical physiology research in which Enroth-Cugell and Robson (1966) determined the spatio-temporal characteristics of X and Y retinal ganglion cells. They observed a sustained response to high-resolution stimuli in X cells, but a transient response to low-resolution stimuli in Y cells. Hubel and Wiesel (1959, 1962) found that this dichotomy was preserved at the lateral geniculate nucleus: Y cells dealing with a transient, gross analysis of the stimulus project 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. Computational vision theorists picked up on this temporal and anatomical distinction to derive models of early visual processes, including edge extraction, stereopsis and motion (see Marr, 1982, for discussions and examples).

In recognition, researchers soon realized that algorithms could not operate on raw pixel values from a digitized picture. A multiscale representation of the image was required 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 notoriously noisy and represent confusing details that would be absent from a coarser representation. Fine scale details, however, are often required when the objects to be distinguished are similar, or more generally, when the task requires detailed information. An efficient strategy may initially produce a stable, but coarse, description of the image before the noisier, but finer, information is extracted for successful categorization. In other words, the LSFs may be extracted and used before the HSFs. This is the *coarse-to-fine hypothesis*.

The notion of a coarse-to-fine recognition strategy is more often assumed than explicitly stated. Parker and Costen (1999, p. 18) eloquently summarize the general view: "the lower spatial frequencies in an image are processed relatively quickly while progressively finer spatial information is processed more slowly." The status of the coarse-to-fine hypothesis remains to be clarified. Is there a physiological bias in the temporal availability of coarse and fine scale information, with LSFs being extracted before HSFs? Would such a bias be 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 another reason, namely that an efficient strategy for recognizing complex images first produces a coarse skeleton of the input which is then fleshed out with fine scale details? (i.e., a strategically driven categorization scheme).

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, Maffei & Sandini, 1983; Parker & Costen, 1999; Parker, Lishman & Hughes, 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 vs. strategical status of this fixed coarse-to-fine scheme was not addressed until recently (see Morrison & Schyns, in press).

Schyns and Oliva (1994) used hybrid stimuli similar to those of Figure 2 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.

For their first experiment, Schyns and Oliva (1994) used a matching task whereby a sample was presented for either 30 ms or 150 ms followed immediately by a mask and then a target. Participants indicated whether or not the sample matched the target. Samples were either full 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, the one depicted in LSFs and the one in HSFs. The two scenes represented by one hybrid could both be matched with their respective target at 30 and 150 ms durations. Nevertheless, exposure duration changed the interpretation of the hybrids: Short exposures elicited more accurate matchings of LSF-hybrids compared with the long exposures (d’ = 2.08 vs. 1.4) , whereas the converse was true of HSF-hybrids (d’ = 1.06 vs. 3.0). 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, 1991). In a second experiment Schyns & 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 ms each with no ISI. An animation contained two distinct sequences, one coarse-to-fine and the other fine-to-coarse--i.e., observers saw two different scene sequences simultaneously. For example, if the top hybrid from Figure 2 is immediately followed by that on the bottom, the coarse-to-fine sequence would represent a motorway while the fine-to-coarse animation would depict a city. It is important to stress that each trial consisted of a single presentation of each of the two hybrids. When asked to name the animated 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 (see also Breitmeyer, 1984; Fiorentini, Maffei & Sandini, 1983; Parker & Costen, 1999; Parker, Lishman & Hughes, 1992, 1997).

Flexible usage hypothesis

An alternative to the fixed coarse-to-fine hypothesis was put forward by Oliva and Schyns (1997; Schyns & Oliva, 1999). The images in figures 1 and 2 can be categorized in a number of ways depending on the use of LSF vs. HSF perceptual cues. In general, the cues subtending different categorizations might themselves be associated with different regions of the spatial spectrum. For example, Schyns and Oliva (1999) showed that the perceptual cues most useful for judging the identity, gender and expression of a face were associated with different spatial scales (see also Sergent, 1986). Thus, the observer who categorizes an image might be biased to the spatial scales with which task relevant perceptual cues are associated. Schyns and Oliva (1999) suggested that rather than being fixed in a coarse-to-fine sequence, the scale usage for categorization could be flexible and determined by the usefulness (or diagnosticity) of cues at different scales. We call this the *flexible usage hypothesis*. In contrast, the coarse-to-fine hypothesis neglects the nature of the categorization task and its information requirements. In the flexible scale usage, the perceptual processing of an identical visual input may be influenced by the nature of the categorization task (Schyns, 1998). There is indeed evidence that categorization can influence the construction of the image percept (e.g., Schyns, Goldstone & Thibaut, 1998).

In a recognition task, Oliva and Schyns (1997, Experiment 1) demonstrated that the LSF and HSF components of a hybrid scene presented for 30 ms both primed subsequent recognition of a full spectrum scene. This indicates that the coarse and fine scale cues are both available early, arguing against a mandatory, perceptually-driven coarse-to-fine scheme. In the related domain of global-to-local processing, researchers have shown that the effect of global precedence was itself modulated by task constraints. For example, Grice, Graham and Boroughs (1983) illustrated 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 Sergent, 1982; and Kimchi, 1992, for a review).

The flexible usage hypothesis suggests that categorization mechanisms tune into the scales that represent information relevant to the task at hand. Two factors need to be considered: the categorization task which specifies the visual information required to resolve this categorization, and the multiple levels of representation of this visual information across the different spatial scales. Flexible use might result from a selective use of only a few of these levels for the task at hand. Oliva and Schyns (1997) and Schyns and Oliva (1999) reported data consistent with the flexible stance. In Oliva and Schyns' (1997) second experiment observers saw scenes, each presented for 135 ms, to identify (*city, highway, living room* or *bedroom*?). They first saw images meaningful at LSFs or HSFs only--e.g., a fine scale highway combined with coarse scale noise. Without discontinuity in presentation, the following images were hybrids--e.g., HSFs depicted a city and LSFs a motorway. Observers identified the hybrids according to the scale at which diagnostic information was initially presented. That is, observers 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. Interestingly, observers claimed to be unaware that two different scenes were present in any one hybrid image, arguing against the possibility that observers first perceived two scenes in hybrids and then decided to report the scene consistent with the sensitization phase. This finding suggests that scale usage is flexible and tunes into the scale at which diagnostic information is represented.

The idea that different categorizations of identical visual inputs (e.g., identity, gender, expressive or not) rely on distinct regions of the spatial spectrum is central to the flexible usage hypothesis. If this is the case (we return to this topic later) then the hypothesis of flexible usage predicts that the perception of identical hybrids should depend on the categorization performed. This question was addressed in Schyns and Oliva (1999) using hybrids derived from the faces of unfamiliar people. For example, a neutral female at HSFs may be superimposed with an angry male at LSFs (see Figure 1a). In Experiment 1, stimuli were presented for 50 ms, and the nature of the categorization was found to moderate stimulus perception. To illustrate, when asked whether the face was expressive or not, observers had a tendency to perceive and to report the fine scale face. However, there was no bias for a gender decision and there was a coarse scale bias when asked to specify the face expression as happy, angry or neutral. Again, observers remained unaware of the presence of two faces in any one image. In sum, perception of identical hybrids was determined by the categorization task, suggesting that categorization processes tune into diagnostic information at specific scales.

In their Experiment 2, Schyns and Oliva (1999) isolated the perceptual byproducts of a categorization task. In phase one, two subject groups applied a different categorization task (expressive or not, vs. which expression) to an identical set of hybrid faces, to induce two orthogonal scale biases (to HSF and LSF, respectively). In phase two, all subjects had 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. For example, when one group preferentially categorized the hybrid of Figure 1a as a female on the basis of its HSF, the other group categorized the same picture as a male on the basis of its LSF. Note that groups only differed on the frequency bandwidth bias acquired in the first phase of the experiment. It is important to stress that 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 markedly differently. From this perceptual transfer we can conclude that categorization can modify the perception of scale information. Note, however, that we established this flexible scale usage at very brief (30 ms) exposures. For longer exposures, and this can be experienced by looking at the hybrids of Figure 1, saccadic eye movements take place and the fine scale seems to dominate perception.

To summarize, it is often assumed, but rarely tested, that spatial scales are processed in a coarse-to-fine manner (e.g., Marr & Hildreth, 1980; Watt, 1987). It would seem that LSFs are extracted before HSFs from simple sinewave 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). However, this does not imply the existence of a mandatory recognition strategy using information from coarse to fine. In fact, the evidence (Oliva & Schyns, 1997; Schyns & Oliva, 1999) conflicts with the view that scale usage for categorization is fixed, and rather suggests it is flexible and driven by the presence of diagnostic information at different scales. Furthermore, converging evidence suggest that the diagnostic use of coarse and fine scale cues in categorization tasks does change the perceptual appearance of the incoming stimulus.

SEARCHING FOR DIAGNOSTIC SCALE INFORMATION

The work reviewed so far demonstrates a flexible attentional control on scale information when this scale is diagnostic (e.g., Oliva & Schyns, 1997). There is also evidence that different categorization tasks tap into different scales of the same stimulus (Schyns & Oliva, 1999). The attentional control of scale use could therefore arise from the information demands of different categorization tasks. This section will explore this hypothesis in detail.

Hybrids can be used to ascertain preferred scale usage (e.g., LSF vs. HSF) from the categorization responses of subjects. They can also reveal the scale that is perceived, and inform on the processing of the neglected scale (e.g., Oliva & Schyns, 1997). However, as a general method to search for the information diagnostic of categorization tasks, hybrids are inherently limited. First, it is difficult to create a hybrid composed of more than two different bandwidths of spatial information while preserving the independent perception of each bandwidth. Consequently, hybrids are restricted to dichotomous searches in scale space (e.g., LSF vs. HSF, or mid-frequencies vs. HSF, and so forth). A second shortcoming is that the method does not locate in the image plane the cues that are diagnostic at a specific scale. To illustrate, suppose you recognized the faces of Figure 1 on the basis of HSF cues (e.g., using their eyelashes and the corner of the mouth). From your categorizations, we would know that you preferred to use the HSF plane, but not which cues you used in this plane. To summarize, the search space for diagnostic cues is three-dimensional (the two-dimensional image x multiple spatial scales). Of this space, the hybrid methodology can only search one dimension (the spatial scales), using a dichotomy (e.g., LSF vs. HSF).

[PHIL: je mettrais un alinéa ici]We now turn to a method, called *Bubbles* [PHIL: il faut mettre "a" à cause de SLIP--j'ai changé partout](Gosselin & Schyns, 2001a) that addresses these two shortcomings and therefore generalizes the search for diagnostic cues to the entire three-dimensional space.

The diagnostic information of Identity, Gender, Expressive or Not

In a nutshell *Bubbles* can determine the use of information specific to a categorization task. *Bubbles* samples an input space (here, the 3D space discussed above) to present sparse versions of the stimuli (here, faces). Observers categorize the sparse stimuli (here, into their identity, gender, and expressive or not) and *Bubbles* keeps track of the information samples that [PHIL: n'est-ce pas "led" plutôt?]lead to correct and incorrect categorizations. From this information, *Bubbles* determines how each region of the input space is selectively used in each categorization task, and depicts the selective use with an *effective stimulus*. The following sections discuss in detail the results of [PHIL: première citation, donc tous les noms doivent apparaître]Schyns, Bonnar and

Gosselin (in press, see also Gosselin & Schyns, 2001a, [PHIL: j'ai ajouté une ref. à RAP] and Gosselin & Schyns, 2002).

In this application, the image generation space comprised the two dimensions of the image plane and the third dimension of spatial scales. To compute an experimental stimulus, we decomposed a face picture (see Figure 3a) into 6 bands of non overlapping spatial frequencies of one octave each—with cutoffs at 90, 45, 22.5, 11.25, 5.62, 2.81 cycles per face, from fine to coarse, respectively, see Figure 3b. The coarsest (i.e., 6^{th}) band served as a constant background because it does not contain useful face information, and so only five bandwidth are represented in Figure 3b. We sampled this image space with [PHIL: j'ai mis "bubbles" en italic] *bubbles* of information (hence the name of the technique). The bubbles were a number of [PHIL: je préfère comme suit] Gaussian windows applied to each of the five spatial frequency bands (the size of each bubble was adjusted so that 3 cycles per face was revealed at each scale—i.e., standard deviations of bubbles were [PHIL: ca a plus de sens de parler de "cycles per face" pour ce qui est des tailles des bulles parce que nous parlons de "cycles per face" pour les bandes de fréquences plus haut. Je n'ai rien changé, mais te suggère de le faire.] .13, .27, .54, 1.08, and 2.15 deg of visual angle, from fine to coarse scales, see Figure 3c). Across trials, the locations of all bubbles changed randomly. Thus, after many trials, bubbles sample the entire image space and the search for diagnostic cues is [PHIL: je dirais plutôt : "non-biased" que "exhaustive"]exhaustive.

INSERT FIGURE 3 ABOUT HERE

In a trial, we added the information samples produced by multiplying the scale-specific face information (Figure 3b) with its respective bubbles (Figure 3c) to produce a sparse stimulus (Figure 3e). The subspace revealed by the bubbles was adjusted to maintain categorization of the sparse faces at a set criterion (here, 75% correct). To respond, observers pressed the appropriate keyboard key (i.e., male vs. female; expressive vs. non-expressive; or, e.g., "John").

On any given trial, we can hypothesize that a correct response means that the information samples revealed enough information to categorize the stimulus. An incorrect response means that there was not enough face information in the samples. Across trials, the interaction between the random bubbles and the observer is therefore a random search for diagnostic task information, using the observer to tease apart the information samples into diagnostic and nondiagnostic.

Specifically, we keep track of the locations of the bubbles that lead to correct categorizations in a different CorrectPlane for each scale (henceforth, CorrectPlane(scale), for scale = 1 to 5, from fine to coarse). In each of these planes, we literally added the masks of bubbles (see Figure 3c, for examples of masks) leading to correct responses. In contrast, TotalPlane(scale) is the sum of all masks leading to correct *and* incorrect categorizations.

From CorrectPlane(scale) and TotalPlane(scale), we derive ProportionPlane(scale) = CorrectPlane(scale) / TotalPlane(scale) per observer. ProportionPlane(scale) is the ratio of the number of times a specific region of the input space has led to a successful categorization over the number of times this region has been presented in the experiment. Across subjects, the averaged ProportionPlane(scale) weighs the importance of the regions of each scale for the categorization task at hand (Gosselin & Schyns, 2001a). If all regions were equally important, ProportionPlane(scale) would be uniform. In other words, the probability that any randomly chosen bubble led to a correct categorization of the input would be equal to the expected performance criterion–here, .75. By the same reasoning, regions above (vs. below) the criterion are more (vs. less) diagnostic for these tasks.

We construct a confidence interval around the mean of the ProportionPlane(scale), for each proportion ($p < .01$). Significance is summarized in a DiagnosticPlane(scale) that represents with a 1 a diagnostic proportion and with a 0 a nondiagnostic proportion. The DiagnosticPlane(scale) is a mask that can filter out the nondiagnostic information at each scale of the face image. We can use them to depict the selective use of information in each task. The resulting *effective stimulus* is simply obtained by multiplying the face information at each scale in Figure 3b with the corresponding DiagnosticPlane(scale).

Figure 4 compares the relative use of scale information in the identity (top), gender (middle) and expressive or not (bottom) tasks. The figure reveals a differential use of information across tasks and scales. Whereas the mouth is represented at all scales in identity and expressive it is does not appear at the finest scales in gender. Similarly, the eyes are both represented at all scales in identity, but only one of them figures in the effective face of gender, and both are neglected in the effective face of expressive. The chin is only well defined in identity. Compared to the mouth and the eyes, the nose is much less represented in all tasks.

INSERT FIGURE 4 ABOUT HERE

We can quantify the use of each spatial scale across tasks. To this end, we divided the diagnostic areas revealed at each scale by the total area covered by the face in the image plane. In Figure 4, the histograms represent the use of diagnostic information at different spatial scales–1 means finest, and 4 coarsest scale. The small face pictures illustrate which cues are used. The use of fine scale information (labeled 1 in the histograms, and depicted in the leftmost small picture) differs considerably across tasks. It depicts the eyes, the mouth and the chin in identity, whereas in gender the finest scale is only used for the left side eye, and in expressive for the mouth. In contrast, the coarsest scale (i.e., the fourth scale) is much less differentiated. It forms a skeleton that is progressively fleshed out with increasing spatial resolution (see the progression of face information from coarse to fine in the small pictures of Figure 4, from right to left.)

In sum, *Bubbles* can search for the information relevant for different categorizations of the same stimuli. It extends the hybrid method presented earlier because it can pin-point the exact location of the diagnostic features in a complex image space.

Second order holistic features

It is widely accepted that face processing may rely on both componential cues (i.e., local features such as the mouth, nose, eyes, a mole) and noncomponential information (the spatial relations between these features), though how these cues are integrated remains unclear (e.g., Bartlett & Searcy, 1993; Calder, Young, Keane & Dean, 1999; Farah, Wilson, Drain & Tanaka, 1998; Macho & Leder, 1998; see also chapters of Bartlett, Searcy and Abdi; Farah & Tanaka; Murray, Rhodes & Schuchinsky, in this volume for disucssions of configural vs. featural information). 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 as, 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 be associated with different spatial scales. Indeed, Sergent (1986 p. 23-24) argued "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.

Our analysis with *Bubbles* has focussed on information of a strictly componential nature (i.e., each proportion of the ProportionPlanes). When several proportions form a continuous region (as was the case for the diagnostic masks, see Figure 3b), it is tempting to assume that the face features within the regions are themselves used holistically (configurally). However, this is not necessarily the case. For example, an observer could use holistically two nonadjacent areas of the face (e.g., the two eyes, or one eye and the mouth). Conversely, two adjacent components could be used independently, but assigned to the same diagnostic region.

We define a holistic use of information as a conjunctive use of information. Operationally, a holistic use of information implies that the presentation of information from several separate bubbles (a conjunction of information) does drive recognition performance.

Here, we limited the conjunction to two distinct bubbles of information, a second order analysis. Thus, in this section, we perform a second order search for the better conjunctions of features in the gender and the identity tasks of Schyns et al. (in press). We restrict our analysis to five main areas of the faces, known from our experiments to be particularly diagnostic: the left eye, the right eye, the nose, the left portion of the mouth, and the right portion of the mouth. We derive the proportion of correct categorizations associated with all possible conjunctions of the five areas of interest. This is summarized in a 5 x 5 matrix per scale, where each cell represents a feature conjunction. In this cell, we increment a counter each time the stimulus comprised at least one bubble in each of two regions concerned and the categorization was correct. We increment a separate counter every time the conjunction was presented, irrespective of response. We perform this simple analysis for all stimuli, subjects and responses[1], and the resulting proportion correct is the division of the two counters. We then compute the significant proportions. Figure 5 depicts the DiagnosticPlane(scale) for feature conjunctions (the plane is now a 5 x 5 symmetrical matrix).

INSERT FIGURES 5 ABOUT HERE

A white square in Figure 5 indicates a significant feature conjunction at one of the scales. The DiagnosticPlane(scale) are symmetrical, but to facilitate reading, we have only kept upper triangle of the symmetrical matrices. These results are best interpreted with the potent information depicted in Figure 4. Remember that the first order analysis revealed the importance of the eyes and the mouth to identify faces. Note that the diagnostic conjunctions for IDENTITY involve mostly relationships between the two eyes and the mouth. In GENDER, these relationships involve mostly the left eye (see Figure 4) and both corners of the mouth (in the first and second scales), and a recurrent relationships between the left corner of the mouth and the nose across all scales. Thus, the second order analysis confirms a differentiated use of information across tasks, adopting the form of diagnostic feature conjunctions.

In sum, *Bubbles* is a search for diagnostic information in any *n*-dimensional image generation space, even if the space is abstract. *Bubbles* is therefore *not* restricted to the 2D image plane, but eye scans are. Because bubbles of information are independent samples in the input space, we can compute how second-order relationships between the samples (but also relationship between more than two distinct bubbles of information) contribute to recognition, and thereby ascertain the amount of holistic processing at different scales.

In the reported data, the DiagnosticPlanes were averaged across subjects, but we could have performed the analysis on a subject per subject basis (opening promising research avenues in visual development, the acquisition of perceptual expertise and visual agnosia). The analysis can also be performed by item (i.e., stimulus), to enable a finer understanding of the recognition of each stimulus in well-specified task contexts. Succinctly stated, *Bubbles* is a reverse projection of the memory representation of an object onto the input information. Suitably applied, it could predict, from use of diagnostic information, how early visual filters at different spatial scales would become tuned to optimize the intake of low-level visual information (e.g, contrast and orientation) in different recognition tasks.

GENERAL DISCUSSION

Researchers in face, object and scene recognition are often concerned with questions about object representations. For example, they ask key questions such as: "Are face, object and scene representations viewpoint-dependent (Bülthoff & Edelman, 1992; Hill, Schyns & Akamatsu, 1997; Perrett, Oram & Ashbridge, 1998; Simons & Wang, 1998; Tarr & Pinker, 1989; Troje & Bülthoff, 1996; among many others) ? " "Are these representations holistic (e.g, *view-based*, Poggio & Edelman, 1990; Tarr & Pinker, 1991), or made of smaller components? (e.g, *geons*, Biederman, 1987; Biederman & Cooper, 1991)"; "Are internal representations complete (e.g., Cutzu & Edelman, 1996), or sparse (Archambault, O'Donnell & Schyns, 1999; Rensink, O'Regan & Clark, 1997); two- or three-dimensional (Liu, Knill & Kersten, 1995); colored or not (Oliva & Schyns, 2000; Tanaka & Presnell, 1999) ?" "Are they hierarchically organized in memory (Brown, 1958; Rosch, Mervis, Gray, Johnson & Boyes-Braem, 1976) ?" "If so, is there a fixed entry point into the hierarchy (Gosselin & Schyns, 2001b; Jolicoeur, Gluck & Kosslyn, 1984; Tanaka & Taylor, 1991) ?" "What is the format of memory representations, and does it change uniformly across the levels of a hierarchy?" (Jolicoeur, 1990). "Does expertise modify memory representations (Tanaka & Gauthier, 1998; Schyns & Rodet, 1997) and the entry point to recognition (Tanaka & Taylor, 1991) ?"

To address these issues, researchers should embrace powerful methodologies that can assign the credit of behavioral performance (e.g., viewpoint-dependence, configural effects, color, speed of categorization, point of entry, expertise effects and so forth) to properties of face, object and scene representations in memory. However, the relationship between behavior and representations is tenuous, making representational issues the most difficult to approach experimentally.

In this chapter, we have taken an alternative approach that allows a rigorous understanding of the recognition process, without asking direct questions about unobservable memory representations. Our analysis builds on the *selective use of diagnostic information*, an important but neglected component of recognition. People who recognize faces, objects and scenes do not use all the information available to them, but instead select the most useful (i.e., diagnostic) elements for the task at hand. The visual system knows what this information is, and how it should be selectively extracted from the visual array to perform flexible categorizations of the same input.

To analyze the flexible use of information, we started from a set of plausible building blocks, the output of spatial filters in early vision (spatial scales), and examined how they were used during the recognition process. We explained that distinct visual cues for recognition often reside at different spatial scales, themselves processed by different frequency-specific channels in early vision. We showed 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. Using *Bubbles*, a more powerful methodology, we pin-pointed the scale information responsible for different categorizations of the same face. This is a rigorous depiction (see Schyns et al., 2001 for further formal developments) that opens up a number of new exciting research avenues to bridge the gap between high- and low-level vision.

Attention and perception

The reviewed experiments with hybrids and *Bubbles* demonstrated that attention can exert a selective control on the scale information used for categorization. Further evidence of selective and task-dependent processing can be found in psychophysics. The detection of sinusoidal gratings worsens when spatial frequency varies across trials compared with the same gratings presented in blocks of constant spatial frequency (e.g., Davis & Graham, 1981), consistent with selective activation or monitoring of spatial frequency channels (Hübner, 1996).

The common underpinnings between the hybrid methodology, *Bubbles* and the psychophysics of early vision provide one promising research avenue to specify the 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 to understand whether attention to a diagnostic spatial scale (or neglect of a scale) affects the filtering properties of the earliest stages of visual processing--e.g., contrast thresholds, frequency tuning, orientation selectivity.

In a recent study, Sowden and Schyns (2000) have examined 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 low and high spatial frequency gratings—cued with a distinct tone. They reported a decrement in grating detection when observers were miscued (e.g., when the LSF tone was followed by a HSF grating), supporting the occurrence of an expectancy effect. The categorization task could likewise cue people to scale-specific face, object and scene features. The cueing in Sowden and Schyns (2000) suggests one possible implementation of the categorization-dependent perceptions reported in hybrids: [PHIL: veux-tu dire "sensitivity modulation" plutôt? Sinon, je ne comprends pas.] 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. In this context, *Bubbles* delivers precious information (see Figure 4). It predicts how scale information should be used in different parts of the visual field, for different categorizations of the same stimulus. It is in principle possible to examine how different parts of the visual field become sensitized to contrast and orientation as a function of categorization tasks. Evidence that categorization tasks can exert such early influence would have far reaching implications for classical issues in cognitive science ranging from the depth of feedback loops in early vision, the early vs. late selection models of attention (Pashler, 1998), the bi-directionality of cognition (Schyns, 1998), the sparse vs. 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, 2000 [PHIL: n'apparait pas dans les références comme tel.]), two groups of observers were initially sensitized to identify the faces of famous people at either low or high spatial frequencies (the other scale was noise). After a few trials, and without participants being told of a change, hybrids were presented which 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: 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 as observers sensitized to HSFs detected the face at the unattended scale (for them, the coarse scale face) more accurately than those in 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, though cues at the other scale may permit other categorizations such as detection. 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.

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 compared with smaller versions (which comprise only the coarse scale information of the larger image). For example, using faces again, certain judgments of expressions (e.g. happiness) are more resilient to changes in viewing distance than others (see Jenkins et al., 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, and Boyes-Braem, 1976; for a review, see [PHIL: c'est paru ça--je l'ai changé dans les refs.] Gosselin & Schyns, 2001b). 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. 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 subordinate categorizations, then 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 3D 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 category (e.g., Humpback whale). Object pairs could appear in any one of 5 sizes, corresponding to 12, 6, 3, 1.5, .75 and .38 degrees of visual angle on the screen. Note 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 basic judgments. Their second experiment confirmed the results in a straightforward naming task. Thus, constraints on the 2D proximal projection of 3D distal objects differentially modify the availability of scale-specific information for basic and subordinate categorizations.

In the flexible usage scenario, the requirements of information needed for different categorization tasks determine a bias to the scale where 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 extraction of visual information at different spatial scales (coarse vs. fine) and spatial extents (global vs. local).

CONCLUDING REMARKS

Our main epistemological point is that one can acquire knowledge about the recognition process by carefully studying diagnostic information without asking questions (or even making assumptions) about memory representations (see also Schyns, 1998). This is a powerful approach because the information used encompasses all the visual features that mediate the recognition task at hand. These features therefore have a dual role. For high-level vision, they reflect the information required from memory to categorize the stimulus, and the job of low-level vision is to extract them from the visual array. Succinctly stated, the features involved in a recognition task bridge the gap between memory and the visual array. They set an agenda for research in high- and low-level vision.

End Notes

1. Because few bubbles were presented together at the coarsest and next to coarsest scales, co-occurrences of bubbles were rare and we restrict our analysis to the three finest scales.

Acknolwedegments

The research described here was partially supported by ESRC grants R000237412 and R000223179 to Philippe G. Schyns.

References

Archambault, A., O'Donnell, C., & Schyns, P.G. (1999). Blind to object changes: When learning one object at different levels of categorization modifies its perception. *Psychological Science*, **10**, 249-255.

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

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

Bayer, H.M., Schwartz, O., & Pelli, D. (1998). Recognizing facial expressions efficiently. *IOVS*, **39**, S172.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, **94**, 115-147.

Biederman, I., & Cooper, E.E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology*, **23,** 393-419.

Blakemore, 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 (London)*, 203, 237-260.

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

Brown, R. (1958). How shall a thing be called? *Psychological Review, 65*, 14-21.

Bülthoff, H.H., & Edelman, S. (1992). Psychophysical support for a two-dimensional view theory of object recognition. *Proceedings of the National Academy of Science USA*, **89**, 60-64.

Burt, D.M. & Perrett, D.I. (1997). Perceptual asymmetries in judgements of facial attractiveness, age, gender, speech and expression. *Neuropsychologia,* **35**, 685-693.

Calder, A.J., Young, A.W., Keane, J., Dean, M. (2000). Configural information in facial expression perception. *Journal of Experimental Psychology: Human Perception and 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.

Cutzu, F., & Edelman, S. (1996). Faithful representations of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Science*, **93**, 12046-12050.

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*. Oxford University Press: New York.

Enroth-Cugel, C., & Robson, J.D. (1966). The contrast sensitivity of retinal ganglion cells of the cat. *Journal of Physiology (London)*, **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 sacadic eye-movements emphasizes visual boundaries. *Perception and 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. (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 and J. P. Thomas (Eds.). *Handbook of Perception and Human Performance, II: Cognitive Processes and Performance*. NY: 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. (2002). RAP: a new framework for visual categorization. *Trends in Cognitive Science. 6,* 70-77.

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

Gosselin, F. & Schyns, P. G. (2001b). Why do we SLIP to the basic-level? computational constraints and their implementation. *Psychological Review, 108*, 735-758.

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: Lawrence Erlbaum.

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

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.

Hill, H., Schyns, P.G., & Akamatsu, S. (1997). Information and viewpoint dependence in face recognition. *Cognition.* **62**, 201-222.

Hochberg, J. (1982). How big is a stimulus. In J. Beck (Ed.). *Organization and Representation in Perception,* 191-217. Lawrence Erlbaum, NJ.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat's striate cortex. *Journal of Physiology*, **148**, 574-591.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex. *Journal of Physiology*, **160**, 106-154.

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

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

Jolicoeur, P. (1990). Identification of disoriented objects: A dual-systems theory. *Mind and Language*, **5**, 387-410.

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

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

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.

Liu, Z., Knill, D.C., & Kersten, D. (1995). Object classification for human and ideal observers. *Vision Research*, **35**, 549-568.

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 and Performance*, **24**, 1486-1500.

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.

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

Morrisson, D. & Schyns, P. G. (1998). Exploring the interactions between face processing and attention. *Perception*, **27**, 130.

Morrison, D.J. & Schyns, P.G. (1999). Spatial scales in the Margaret Thatcher Illusion. Submitted for publication..

Morrison, D.J. & Schyns, P.G. (in press). Usage of spatial scales for the categorization of faces, objects and scenes. *Psychological Bulletin and Review.*

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

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

Perrett, D.I., Oram, M.W., & Ashbridge, E. (1998). Evidence accumulation in cell populations responsive to faces: an account of generalisation of recognition without mental transformation. *Cognition*, **67**, 111-145.

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.

Oliva, A. & Schyns, P.G. (2000). Diagnostic colors mediates scene recognition. *Cognitive Psychology*, **41**, 176-210.

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. (1997). Evidence for the view that temporospatial integration in vision is temporally anisotropic. *Perception*, **26**, 1169-1180.

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

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

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

Rensink, R.A., O'Regan, J.K., & Clark, J.J. (1997). To see or not to see: the need for attention to perceive changes in scenes. *Psychological Science,* **8**, 368-373.

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. & 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.

Schyns, P.G., & Rodet, L. (1997). Categorization creates functional features. *Journal of Experimental Psychology: Learning, Memory & Cognition,* **23**, 681-696.

Schyns, P. G., Bonnar, L. & Gosselin, F. (in press). Show me the features! Understanding Recognition from the Use of Visual Information. *Psychological Science*.

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

Sergent, J. (1982). Theoretical and methodological consequences of variations in exposure duration in visual laterality studies. *Perception and 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.

Simons, D., & Wang, R.F. (1998). Perceiving real-world viewpoint changes. *Psychological Science.* **9**, 315-320.

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

Sowden, 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**, 34.

Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. In R.L. Goldstone, D.L. Medin, & P.G. Schyns (Eds.), *Perceptual Learning*. San Diego: Academic Press.

Tanaka, J.W., & Presnell, L.M. (1999). Color diagnosticity in object recognition. *Perception & Psychophysics*, **61**, 1140-1153.

Tanaka, J., & Taylor, M.E. (1991). Object categories and expertise: Is the basic level in the eye of the beholder? *Cognitive Psychology,* **15**, 121-149.

Tarr, M.J., & Pinker, S. (1989). Mental rotation and orientation-dependence in shape recognition. *Cognitive Psychology*, **21**, 233-282.

Tarr, M.J., & Pinker, S. (1991). Orientation-dependent mechanisms in shape recognition: Further issues. *Psychological Science*, **2**, 207-209.

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

Troje, N. & Bülthoff, H.H. (1996) Face recognition under varying pose: The role of texture and shape. *Vision Research*, **36**, 1761-1771.

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.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research*, 25, 1661-1674.

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 and O. Firschein (Eds.), *Readings in Computer Vision: Issues, Problems, Principles and Paradigms*. Morgan Kaufmann.

Figure captions

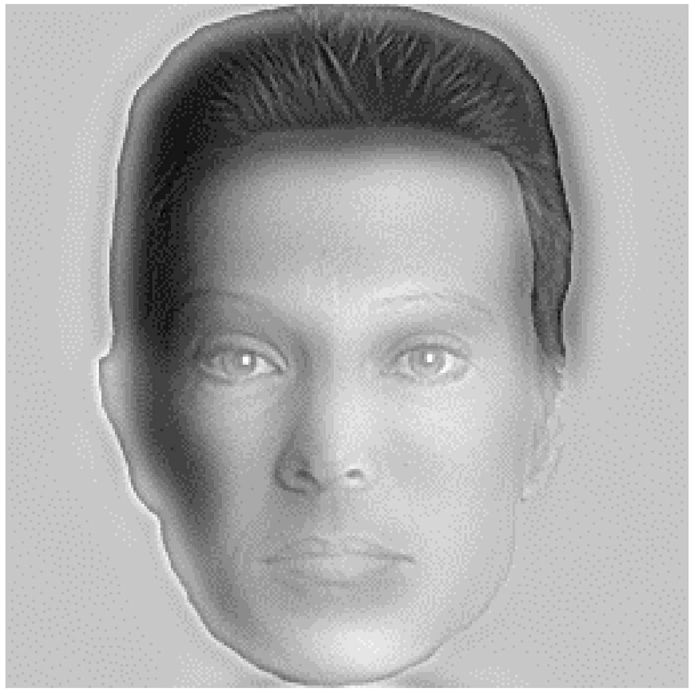
Figure 1. Figure 1 (from Schyns & Oliva, 1999) illustrates two hybrid faces. The fine spatial scale (High Spatial Frequencies, or HSF) represents a nonexpressive woman in the top picture and an angry man in the bottom picture. The coarse spatial scale (Low Spatial Frequencies, or LSF) represents the angry man in the top picture and the neutral woman in the bottom picture. To see the LSF faces, squint, blink, or step back from the picture until your perception changes.

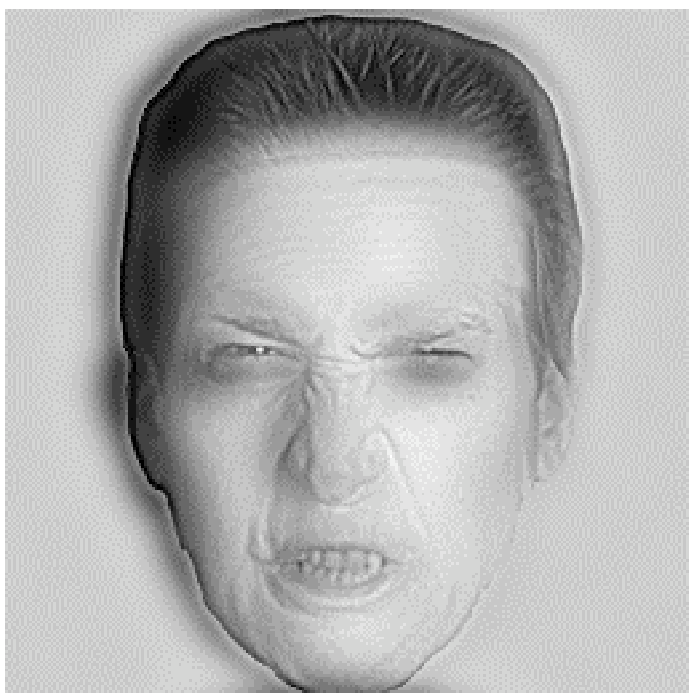
Figure 2. This figure (adapted from Schyns & Oliva, 1994) shows two examples of hybrid scenes. The top picture mixes the fine information of a city with the coarse information of a highway. The bottom picture mixes the opposite information.

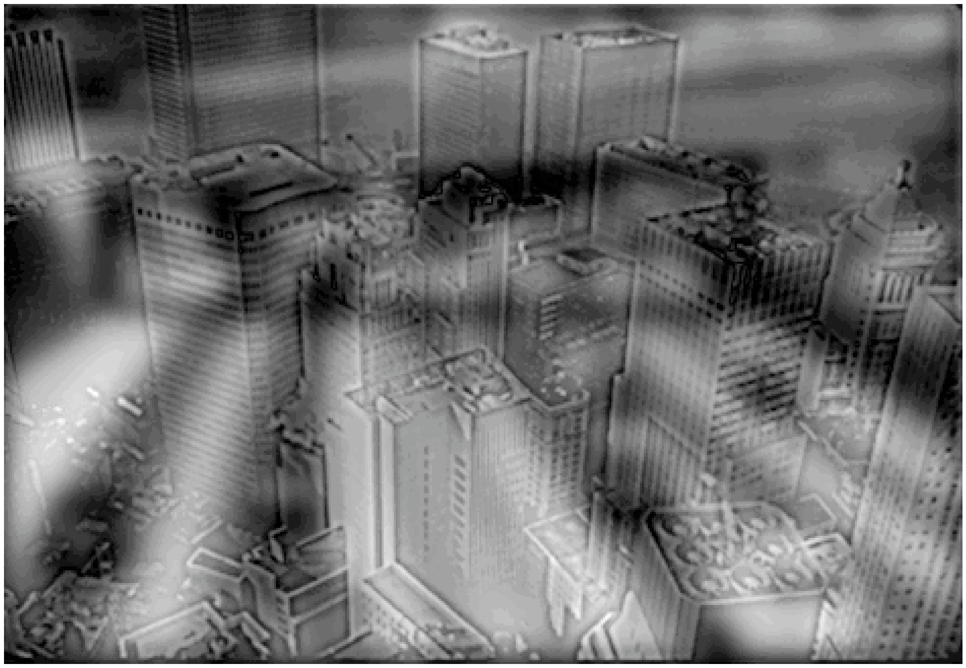
Figure 3 illustrates the application of *Bubbles* to the 3D space composed of a 2D face [PHIL:](adapted from Gosselin & Schyns, 2001a). Pictures in (b) represent five different scales of (a); (c) illustrate the bubbles applied to each scale; (d) depict the information of the scales of (b) sampled by the bubbles of (c). Note that on this trial there is no revealed information at the fifth scale. By integrating the pictures in (d) we obtain (e), a stimulus subjects actually saw.

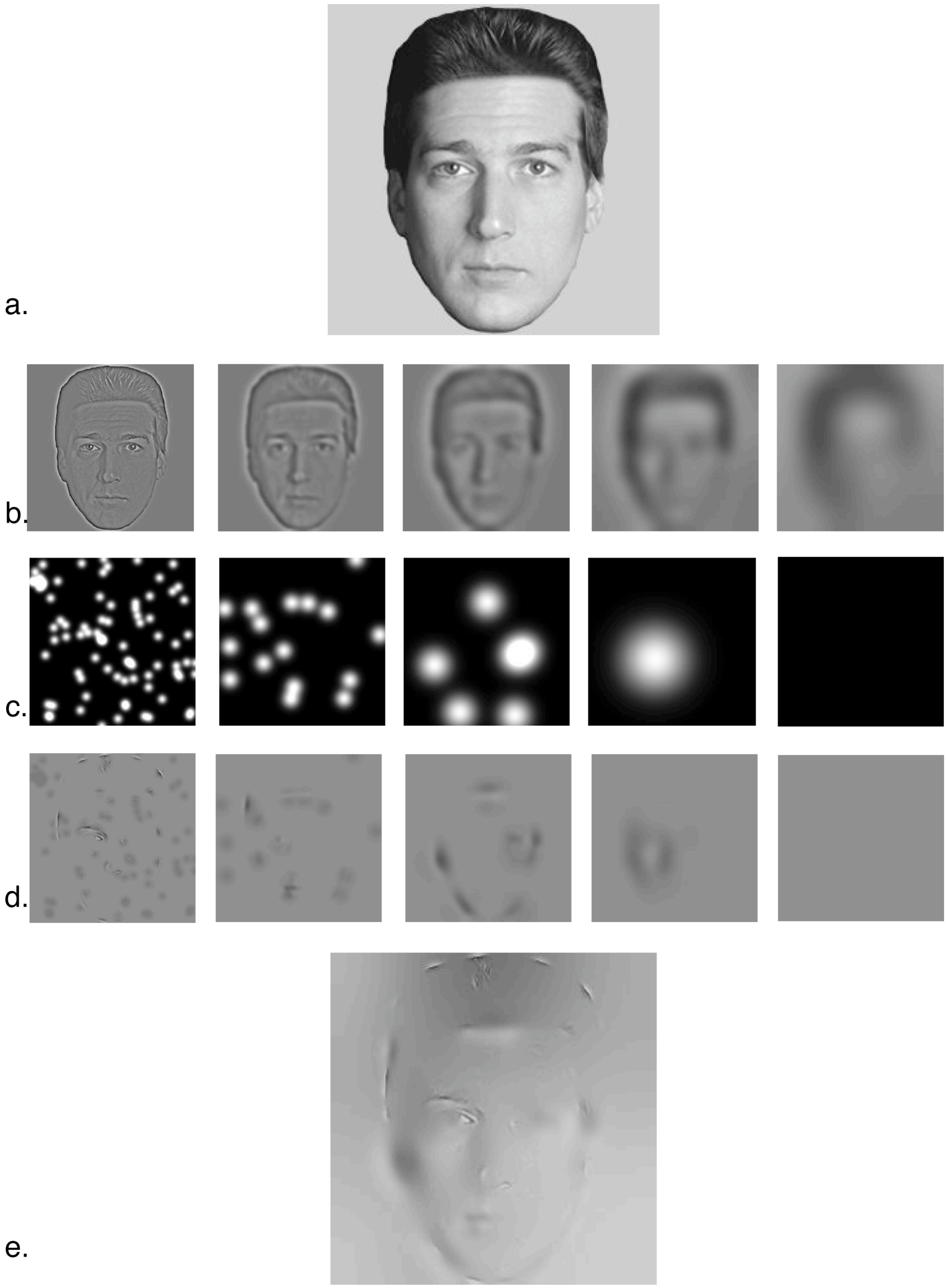
Figure 4. (a) The larger face depicts the effective face stimulus for the identity task [PHIL:] (adapted from Schyns, Bonnar & Gosselin, in press). The smaller pictures illustrate the diagnostic information used to resolve the identity task at each independent scale from fine to coarse, respectively. The coarsest scale is not depicted as it contains no meaningful information. The bar chart provides a quantitative illustration of the proportion of the face area used to resolve the task at each scale. Figures (b) and (c) follow the same format as figure (a) illustrating the potent face for the gender task and expressive or not task respectively, the diagnostic information for each task at each scale and a quantitative account of the use of information in the bar charts.

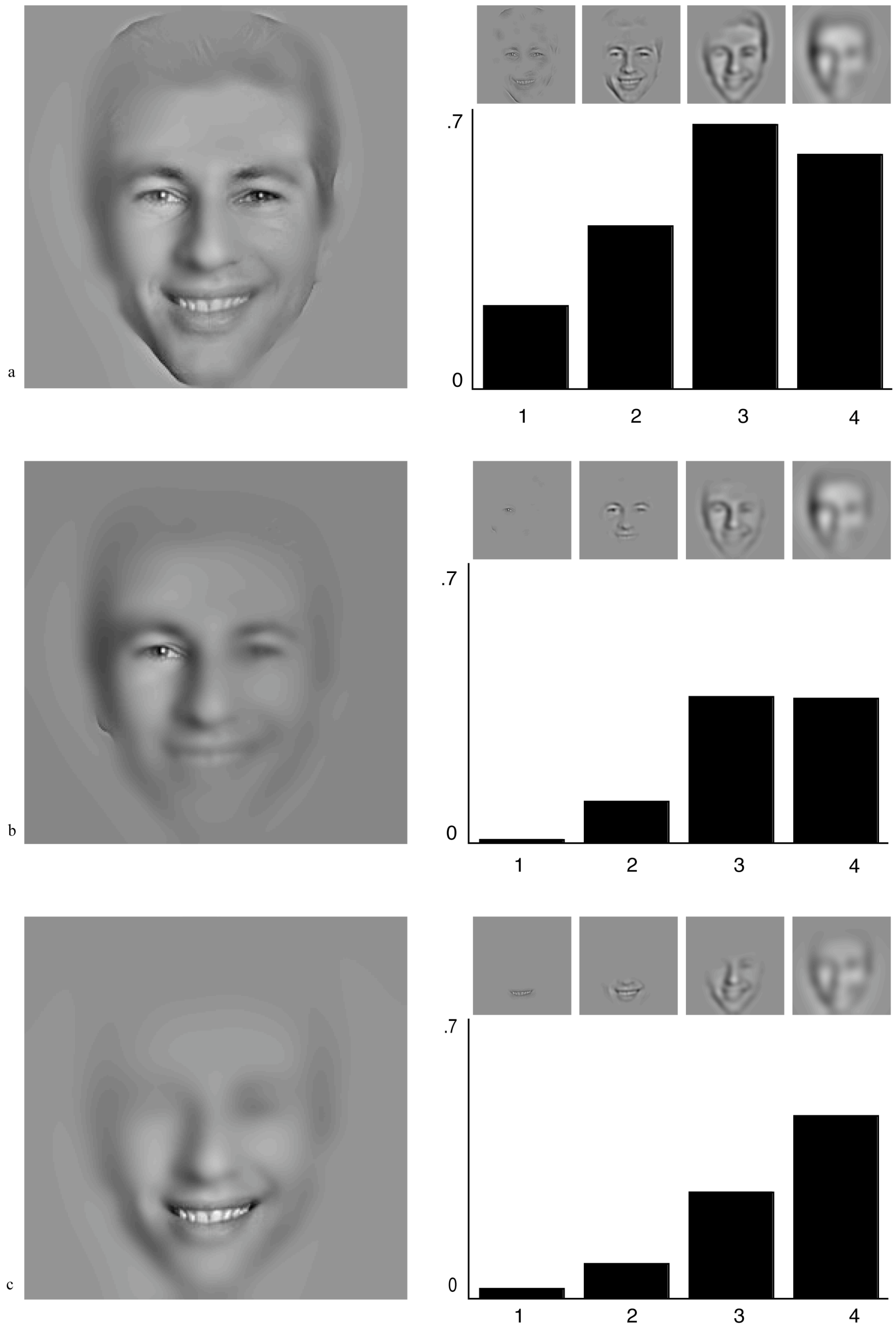
Figure 5. The diagnostic feature conjunctions resulting from the second-order analysis of *Bubbles* for the Identity and Gender tasks [PHIL:] (adapted from Schyns, Bonnar & Gosselin, in press). At each scale, a white square reveals a significant conjunction of features that drove recognition performance. Note that the symmetry of the DiagnosticPlane(scale) has been eliminated to improve the readability of the matrices.



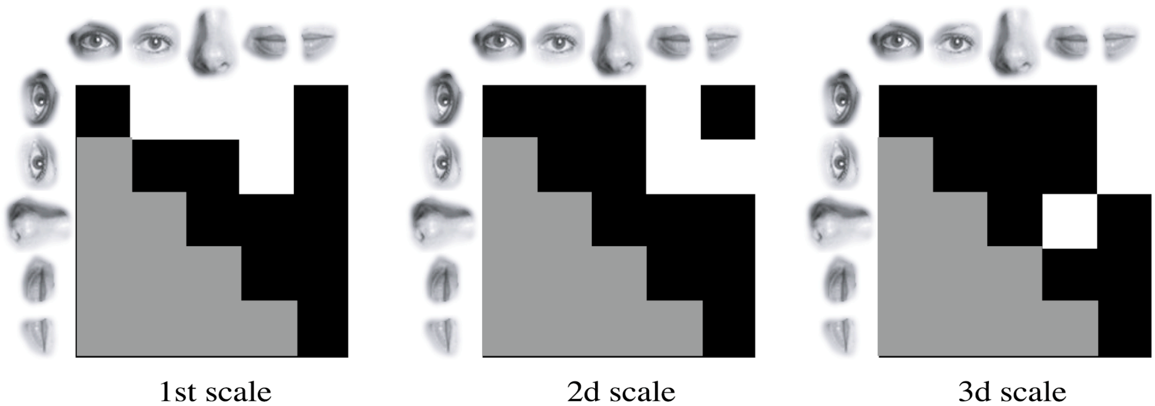





a.
b.
c.
d.
e.


a
.7
0
1
2
3
4
b
.7
0
1
2
3
4
c
.7
0
1
2
3
4

Identity



Gender