Research Article

SHOW ME THE FEATURES! Understanding Recognition From the Use of Visual Information

Philippe G. Schyns,[1] Lizann Bonnar,[1] and Frédéric Gosselin[2]

[1]University of Glasgow, Glasgow, United Kingdom, and [2]University of Montréal, Montréal, Canada

Abstract—*We propose an approach that allows a rigorous understanding of the visual categorization and recognition process without asking direct questions about unobservable memory representations. Our approach builds on the selective use of visual information in recognition and a new method (Bubbles) to depict and measure what this information is. We examine three face-recognition tasks (identity, gender, expressive or not) and establish the componential and holistic information responsible for recognition performance. On the basis of this information, we derive task-specific gradients of probability for the allocation of attention to the different regions of the face.*

In recent years, most face-, object-, and scene-recognition researchers have gathered around a common agenda: to understand the structure of representations in memory. A number of fundamental issues have been articulated, and researchers typically ask questions such as the following: 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)? Are they two- or three-dimensional (Liu, Knill, & Kersten, 1995)? Are they colored or not (Oliva & Schyns, 2000; Tanaka & Presnell, 1999)? Are they hierarchically organized in memory (Brown, 1958; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976)? Is there a fixed entry point into the hierarchy (Gosselin & Schyns, 2001b; Jolicoeur, Gluck, & Kosslyn, 1984; Tanaka & Taylor, 1991)? Does expertise modify memory representations (Tanaka & Gauthier, 1997; Schyns & Rodet, 1997)? What is the entry point to recognition (Tanaka & Taylor, 1991)? What is the format of memory representations, and does it change uniformly across the levels of a hierarchy (Jolicoeur, 1990)?

To address these complex issues, categorization and recognition researchers should be equipped with methodologies of a commensurate power—methodologies that can assign the credit for behavioral performance (e.g., viewpoint dependence, configural effects, color, speed of categorization, point of entry, expertise effects) to specific properties of the representations of visual events in memory. However, the relationship between behavior and representations is tenuous, making representational issues the most difficult to approach experimentally.

In this article, we propose 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 visual events do not use all the information impinging on the retina, but instead use only the elements that are most useful (i.e., diagnostic) for the task at hand. In most instances, this information is not available to conscious experience, but the visual system knows what it is, and how to selectively extract diagnostic information from the visual array to perform multiple categorizations of the same input (e.g., Schyns & Oliva, 1999).

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[1] 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. For low-level vision, they specify which information to extract from the visual array. In short, the features involved in a recognition task bridge between memory and the visual array. They set the agenda for high- and low-level vision. Now, let us see what these features are.

EXPERIMENT

The experiment used Bubbles (Gosselin & Schyns, 2001a) to visualize the information used in three face-categorization tasks. Faces are good stimuli for our demonstrations: Their compactness enables a tight control of presentation that limits the spatial extent of useful cues; the familiarity of their categorization simplifies the experimental procedure. However, the principles developed with faces also apply to other visual events (including objects and scenes).

In a between-subjects design, different subject groups performed three different categorization tasks (identity, gender, expressive or not) on the same set of 10 faces (5 males, 5 females), each displaying two possible expressions (neutral vs. happy). Prior to the experiment, we had all subjects learn the three categorizations of the 10 faces, in order to normalize exposure to and expertise with the stimuli.

We determined the face information used for each task by having Bubbles sample an input space to present sparse versions of the faces as stimuli. The subjects categorized these stimuli, and Bubbles kept track of the samples of information that led to correct and incorrect categorization responses. From this performance information, we were able to establish how each region of the input space is selectively used

Address correspondence to Philippe G. Schyns, Department of Psychology, University of Glasgow, 56, Hillhead St., Glasgow G12 8QB, United Kingdom; e-mail: philippe@psy.gla.ac.uk.

1. Feature is a loaded term in vision science, and its meaning can vary across subdisciplines. Henceforth, we use the term to refer to any elementary property of a distal stimulus that the observer uses to resolve the cognitive task at hand (Schyns, Goldstone, & Thibaut, 1998). We show that even though the method we discuss, Bubbles, does not strictly isolate the representations of these features, it is reasonable to assume that Bubbles reveals the portions of the input space in which specific instances of the diagnostic features are represented.

in each categorization task, and to derive the *effective stimulus*, which depicts this selective use of information.

Method

Participants

Participants were 45 paid University of Glasgow (Glasgow, United Kingdom) students, with normal or corrected-to-normal vision. Each participant was randomly assigned to one of the three experimental groups ($n = 15$ per group), who categorized, respectively, the identity, gender, and expressiveness (expressive or not) of the faces.

Stimuli

The experiment ran on a Macintosh G4 computer using a program written with the Psychophysics and Pyramid Toolboxes for Matlab (Brainard, 1997; Pelli, 1997; Simoncelli, 1997). Stimuli were generated from the gray-scale faces of Schyns and Oliva (1999; 5 males and 5 females, each of whom displayed two different expressions, neutral and happy; hairstyle, global orientation, and lighting were normalized).

We applied Bubbles to an image-generation space composed of three dimensions (the standard *x*- and *y*-axes of the image plane, plus a *z*-axis representing spatial frequencies). We chose to search spatial frequencies because psychophysical studies have established that early vision analyzes the input at multiple scales (see De Valois & De Valois, 1990, for a review).

Figure 1 illustrates the stimulus-generation process. To compute each stimulus, we first decomposed an original face into six independent bands of spatial frequencies of one octave each. The cutoffs of the scales were, respectively, 90, 45, 22.5, 11.25, 5.62, and 2.81 cycles per face, from the finest to the coarsest scale. The coarsest band was a constant background and is not depicted in Figure 1.

The face represented at each band (see Fig. 1b) was then partly revealed by puncturing a number of randomly located Gaussian windows (henceforth called bubbles) in a black mask covering the face area of the image (see Fig. 1c). We normalized to 3 the number of cycles per bubble that any bubble could reveal and adjusted the size of the Gaussian for each frequency band accordingly (standard deviations of the bubbles were 0.13, 0.27, 0.54, 1.08, and 2.15 cycles/deg of visual angle, from the finest to the coarsest scale). Because the size of the bubbles increased as the scale became more coarse, the number of bubbles differed across scales so that the total area of the face revealed, on average, would remain constant. To generate a sparse face (see Fig. 1e), we added together the partial face information revealed at all the scales (see Fig. 1d).

To maintain 75%-correct categorization of sparse faces, we adjusted (on a trial-by-trial basis) the number of bubbles (i.e., the total subspace revealed by the bubbles). In sum, Bubbles performed an asymptotically exhaustive random search of the input space (see Gosselin & Schyns, 2001a, for a detailed discussion of the properties of Bubbles).

Procedure

Prior to the experiment, the participants' experience with the stimuli was normalized by showing them printed pictures with corresponding names at the bottom and requiring them to learn to criterion (perfect identification of all faces twice in a row) the gender, expression, and name attached to each face.

The experiment comprised two sessions of 500 trials (25 presentations of each of the 20 faces), but we analyzed the data from only the last 500 trials, when subjects were really familiar with the faces and experimental procedure. In each trial, one sparse face computed as described earlier appeared on the screen. To respond, participants pressed labeled keys on a computer keyboard. The task was self-paced (but we nevertheless collected reaction times), and no feedback was given. A chin rest maintained subjects at a constant 100-cm viewing distance. Stimuli subtended $5.72° \times 5.72°$ of visual angle on the screen.

In the *identity* condition, participants had to determine the identity of each sparse face (from 10 possibilities). In the *gender* condition, they were instructed to decide whether the stimulus was male or female. In the *expressiveness* condition, they had to judge whether the sparse face was expressive or not (smiling or neutral). The three groups therefore performed different categorizations of the same stimulus set.

Results

On average, subjects required a total of 33, 20, and 15 bubbles to reach the 75%-correct performance criterion in the identity, gender, and expressiveness conditions, respectively. We hypothesize that on any given trial, if the subject could correctly categorize the sparse face on the basis of the information revealed by the bubbles, that information was sufficient for that categorization. Across trials, we therefore kept track of the locations of the bubbles leading to correct categorizations. To this end, for each scale we added the masks with the bubbles leading to correct categorizations in order to create a *CorrectPlane*—henceforth, CorrectPlane(scale), where scale = 1 to 5, from fine to coarse (see Fig. 1c for examples of masks). CorrectPlane(scale) therefore represents, for each scale, the locations where sampling of face information (bubbles) led to correct responses. We also added the masks with bubbles leading to both correct and incorrect categorizations to create *TotalPlane(scale)*. So TotalPlane(scale) represents, for each scale, the total sampling frequency of face information.

From the information in CorrectPlane(scale) and TotalPlane(scale), we determined, for each subject separately and on a cell-by-cell basis, the ratio of the number of times a specific location led to a successful categorization over the number of times this location was presented, CorrectPlane(scale)/TotalPlane(scale). We refer to this ratio as *ProportionPlane(scale)*. Across subjects, the averaged ProportionPlane(scale) weighs the importance of the regions of each scale for the categorization task at hand. If all regions were equally important, ProportionPlane(scale) would be uniform across cells, and equal to the performance criterion—here, .75. Consequently, regions significantly above (vs. below) the performance criterion are more (vs. less) diagnostic of the considered categorization.

To determine this significance, we built a confidence interval ($p <$ .01) around the mean of the ProportionPlane(scale), for each proportion. *DiagnosticPlane(scale)* was created by representing diagnostic (significant) proportions with a filtering weight of 1 and nondiagnostic proportions with a filtering weight of 0. These diagnostic weights were then used to filter the original stimulus to derive the *effective stimulus* (see Fig. 2), which depicts the selective use of information in each task. The effective stimulus is simply obtained by multiplying the face information at each scale in Figure 1b with the corresponding DiagnosticPlane(scale).

Use of scale information in the three categorization tasks

The large face pictures of Figure 2 illustrate the relative use of scale information in the three groups of participants. For example, whereas the mouth is well defined at all scales in the identity and expressiveness conditions, it is neglected at the finest scales in the gender condition. In a related vein, the eyes are both represented at all scales in the identity condition, but only one of them is kept in the gender condition, and both are neglected in the expressiveness condition. The chin is well defined in the identity condition, but not in the other two conditions. Compared with the mouth and the eyes, the nose is less well defined in all tasks.

To quantify the use of spatial scales across tasks, we computed a proportion reflecting the diagnostic face area at each scale (see the histograms and corresponding plots of information in the small face pictures in Fig. 2). In the small face pictures in Figure 2, the use of fine-scale information (i.e., the first scale) is most differentiated across the three tasks. In the identity task, it depicts the eyes, the mouth, and the chin, whereas in the gender task, it is used only for the left eye, and in the expressiveness task, it is used only for the mouth. In contrast, the

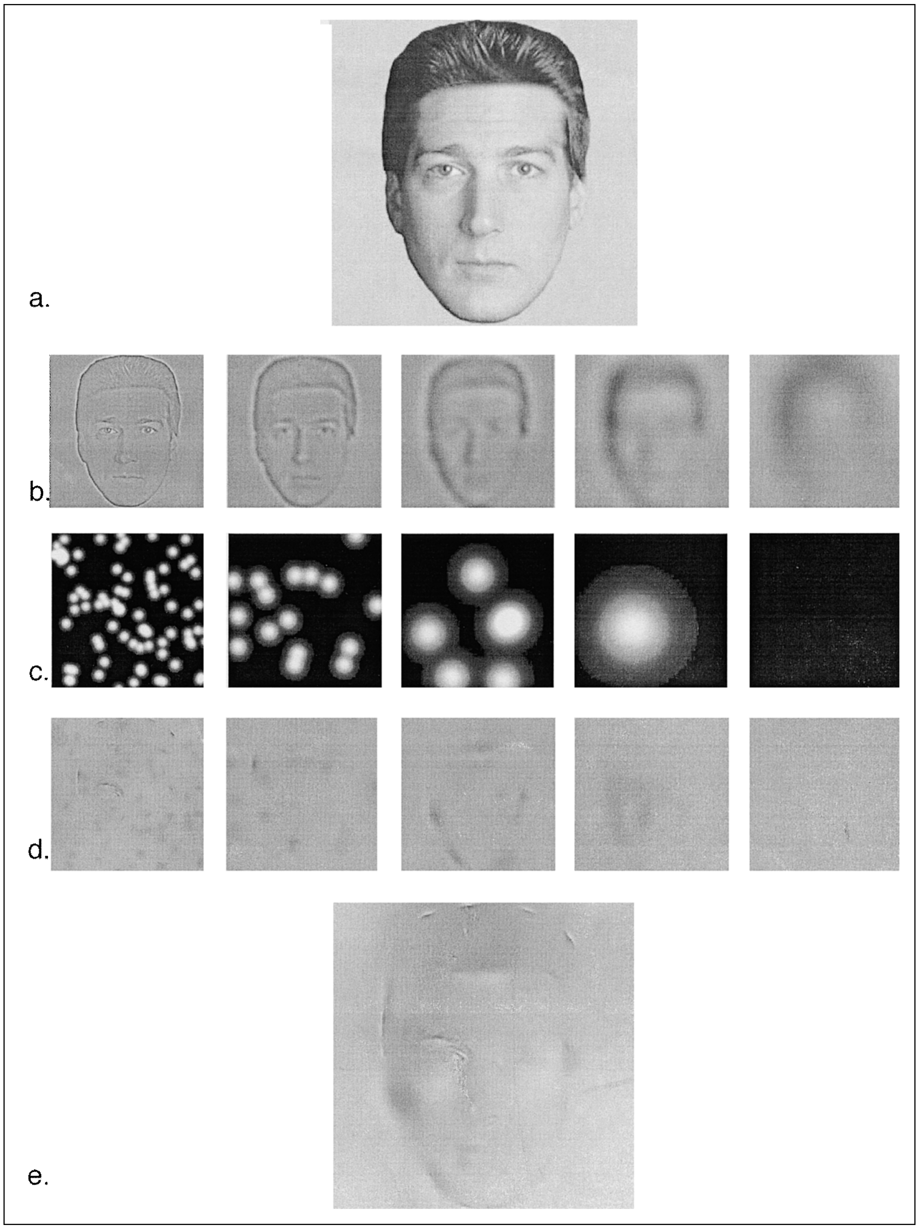

Fig. 1. Application of Bubbles to the three-dimensional space composed of a two-dimensional face and spatial scales on the third dimension. The pictures in (b) represent the information from the face in (a) that is present at five different bands of spatial frequencies (scales). The illustrations in (d) show the information that is revealed at each scale when the face is covered by a mask punctured by the bubbles shown in (c). Note that on this trial there is no information revealed at the fifth scale. The pictures in (d) were integrated to obtain the stimulus subjects actually saw (e).

coarsest scale (i.e., the fourth scale) is much less differentiated across tasks. Thus, the coarse scale forms a skeleton that is progressively distinguished and fleshed out with increasing spatial resolution (see the progression of face information from coarse to fine in the small pictures of Fig. 2, from right to left).

A comparison of the relative use of scales within each task shows a clear advantage in the identity task for the third scale, corresponding to face information between 11.25 and 22.5 cycles per face (the best scale for face recognition varies between 8 and 32 cycles per face in the literature; see Morrison & Schyns, 2001). The preferred scale for

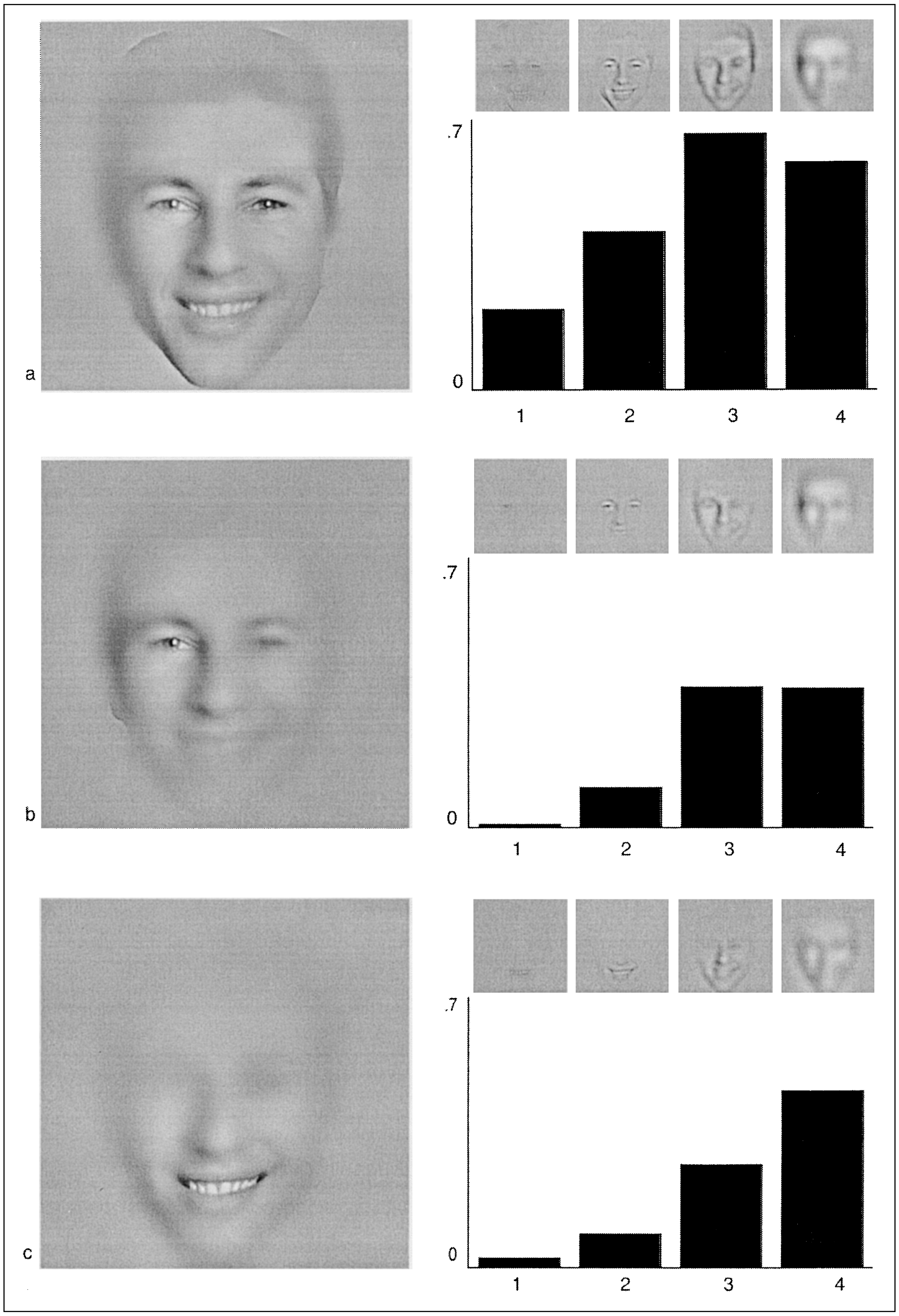

Fig. 2. The effective faces (large faces) and diagnostic information used to resolve the identity (a), gender (b), and expressiveness (c) tasks. The smaller pictures illustrate the diagnostic information used to resolve each task at each independent scale, from fine to coarse, respectively. Results are shown for the first four scales only because there was no meaningful (significant) information at the fifth (coarsest) scale. For each task, the bar graph shows the proportion of the total face area that was used at each scale.

Fig. 3. The two-dimensional attentional maps for the identity, gender, and expressiveness tasks, respectively, from left to right. The gray scale represents the gradient of probability (white = 1, black = 0) of finding diagnostic information at any location of the image.

the expressiveness task (the discrimination between neutral and happy) is information between 5.62 and 11.25 cycles per face (the fourth scale; this result is consistent with Jenkins, Craven, Bruce, & Akamatsu, 1997, and Bayer, Schwartz, & Pelli, 1998). For the gender task, the third and fourth scales are most used. Across tasks, there appears to be a bias for face information between 5.62 and 22.5 cycles per face (the coarser scales) when information is available from the entire scale spectrum.

Deriving a two-dimensional map of attention

We derived a measure of the diagnosticity of each two-dimensional image location from the measurements of diagnostic information at the various spatial scales. The histograms in Figure 2 represent the probability of finding diagnostic information in any of the four diagnostic scales. For each task, we normalized the probabilities represented in each histogram, so that their sum was 1. We then multiplied these probabilities by their corresponding DiagnosticPlane(scale), and added together these partial multiplications, on a pixel-by-pixel basis. This sum represents a gradient of probability of finding diagnostic information at location x, y in the image plane, for a given task. For example, the probability $P(x, y)$ equals 1 when diagnostic information is present at all scales for an image location, and equals 0 when no diagnostic information is present at any scale for the same location.

Figure 3 renders in gray scale the gradient of probability (white = 1, black = 0) of finding diagnostic information at any location of the image in each of the three tasks. If attention is allocated (or eye movements are guided) to the most relevant image locations in a task, the maps of Figure 3 have a predictive value. Even though testing them is beyond the scope of this article, there is evidence that the maps are valid. Figure 2 reveals that for the identity task, the regions of the eyes and the mouth are diagnostic across the entire scale spectrum (see Tanaka & Farah, 1993, for evidence that relationships between these features determine recognition), and so these locations have highest probability in Figure 3 (see eye movement studies from Yarbus, 1965, to Henderson, Falk,

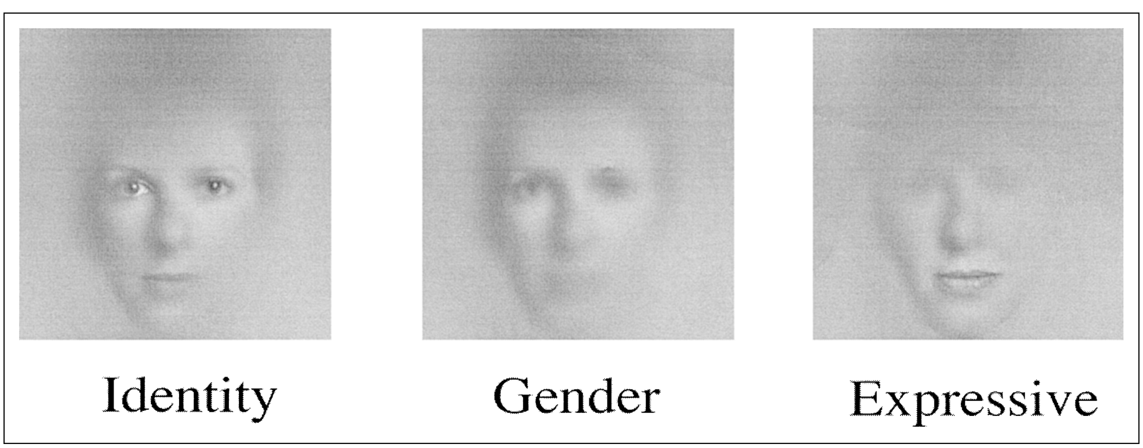

Fig. 4. The face information that discriminated images associated with fast versus slow categorization responses in each of the three tasks.

Minut, Dyer, & Mahadevan, 2001). In Figure 3, the map for the gender task predicts a lateralization to the left side of the image, consistent with the literature (Burt & Perrett, 1997). In sum, attentional maps could provide a rational model of eye movement sequences.

There are two main differences between Bubbles and an eye movement technique. First, Bubbles can search any n-dimensional parameterized image-generation space; it is not restricted to the two-dimensional image plane, as studies of eye movements are. Second, there is an immediate link between a measure of performance and bubbles of information, whereas it is not always clear what observers actually do with the information they fixate.

Reverse-correlating reaction time with Bubbles

If faster categorizations are indicative of a closer match between a category representation and input information, then it is interesting to visualize the information present when subjects responded rapidly, or slowly. We worked backward from the reaction time of observers categorizing sparse faces to the visual information responsible for these categorizations. We first derived the median response time per task (identity: 1.84 s; gender: 0.96 s; expressiveness: 0.97s) and then classified each experimental stimulus according to whether it induced a fast or a slow categorization response (with respect to the median). To create *FastPlane(scale)*, we added together the bubbles of the stimuli leading to the fast categorizations; similarly, to create *SlowPlane(scale)*, we added together the bubbles of the stimuli leading to the slow categorizations. We then derived the discrimination plane[2] between fast and slow categorizations as follows: DiscriminationPlane(scale) = FastPlane(scale) – SlowPlane(scale). The discrimination image thus represents the information of the image space that distinguishes the fast from the slow categorization responses.

Figure 4 illustrates the scale-specific face information that discriminated between fast and slow categorizations of a given face for the identity, gender, and expressiveness tasks. Note that the images in Figure 4 correlate highly with the diagnostic masks revealing the potent faces in Figure 2. This analysis not only has theoretical interest in revealing the high correlation between input information and the information required to categorize the stimulus, but also further validates the diagnostic masks derived earlier.

Higher-order use of information

Our analysis with Bubbles has so far focused on information of a strictly componential nature (i.e., each cell of the ProportionPlanes). However, it is widely accepted that face processing may rely on both componential information (i.e., local features such as the mouth, nose, and eyes) and noncomponential information (the spatial relations between these features), though how these cues are integrated remains unclear (e.g., Farah, Wilson, Drain, & Tanaka, 1998). Here, we show how bubbles of information can be analyzed to reveal the use of higher-order information (i.e., conjunctions of n features; here, $n = 2$).

Operationally, a higher-order (i.e., holistic) use of information leads to better performance when the components are simultaneously presented in distinct bubbles. We restricted our analysis to the five face areas believed to be particularly diagnostic (e.g., Tanaka & Sengco, 1997): the left eye, the right eye, the nose, and the left and right portions of the mouth. A symmetric 5×5 matrix represents all possible conjunctions sampling information from these five regions of interest (see Fig. 5). In each cell of this matrix, we incremented a counter whenever a categorization was correct and the stimulus comprised at least one bubble in each of the two regions concerned. We also incremented a separate counter in the same cell every time the conjunction was presented. We then derived ProportionPlane(scale) by dividing the former counter by the latter counter in each cell. Each proportion thus represents the success of a feature conjunction in the given task. We derived the statistically significant proportions ($p < .01$) to create DiagnosticPlane(scale) for feature conjunctions (see Fig. 5; the plane is now a 5×5 symmetrical matrix).

DiagnosticPlane(scale) is symmetrical, but to facilitate reading, we have kept only the upper triangle of each symmetrical matrix in Figure 5. These results are best interpreted in conjunction with the information in Figure 2. Remember that the first-order analysis revealed the importance of the eyes and the mouth to identify faces. Note that the diagnostic conjunctions for the identity task involve mostly relationships between the two eyes and the mouth. In the gender task, the diagnostic relationships involve mostly the left eye and both corners of the mouth (in the first and second scales), and a recurrent relationship between the left corner of the mouth and the nose across all scales. For the expressiveness task, we found an expected relationship between the two corners of the mouth at the first and third scales, and relationships between the eyes and the corners of the mouth at the first and second scales. Thus, the second-order analysis confirms a differentiated use of information across tasks—in this case, in the form of diagnostic feature conjunctions.

CONCLUDING REMARKS

Our goal was to illustrate a new approach to study recognition without directly asking questions about internal representations. Our analysis established how three face-categorization tasks selectively used componential and configural information from a three-dimensional input space (the two-dimensional image plane $\times$ spatial scales). From this selective use, we derived a gradient of probability of locating diagnostic information in the image plane. A rational human categorizer should selectively allocate attention to the regions of the image that maximize this probability, thus minimizing the uncertainty of locating diagnostic information.

As we argued earlier, diagnostic information should set an agenda for further research in low- and high-level vision. Researchers now have the means of studying the structure of diagnostic information in scale space, and so studies in low-level vision can be carried out to establish how attentional mechanisms intervene to extract this information (e.g., with scale and orientation selectivity). Diagnostic information encompasses all the features of a recognition task, and so studies in high-level vision can be carried out to elucidate what these features actually are. We believe that our approach goes a long way toward resolving the problem of information relevance in visual categorization.

Acknowledgments—This research was supported by Economic & Social Research Council Grant R000223179 to Philippe G. Schyns.

2. DiscriminationPlane(scale) is the best least squares estimate (Ahumada & Lovell, 1971) of the face information leading to fast versus slow response times.

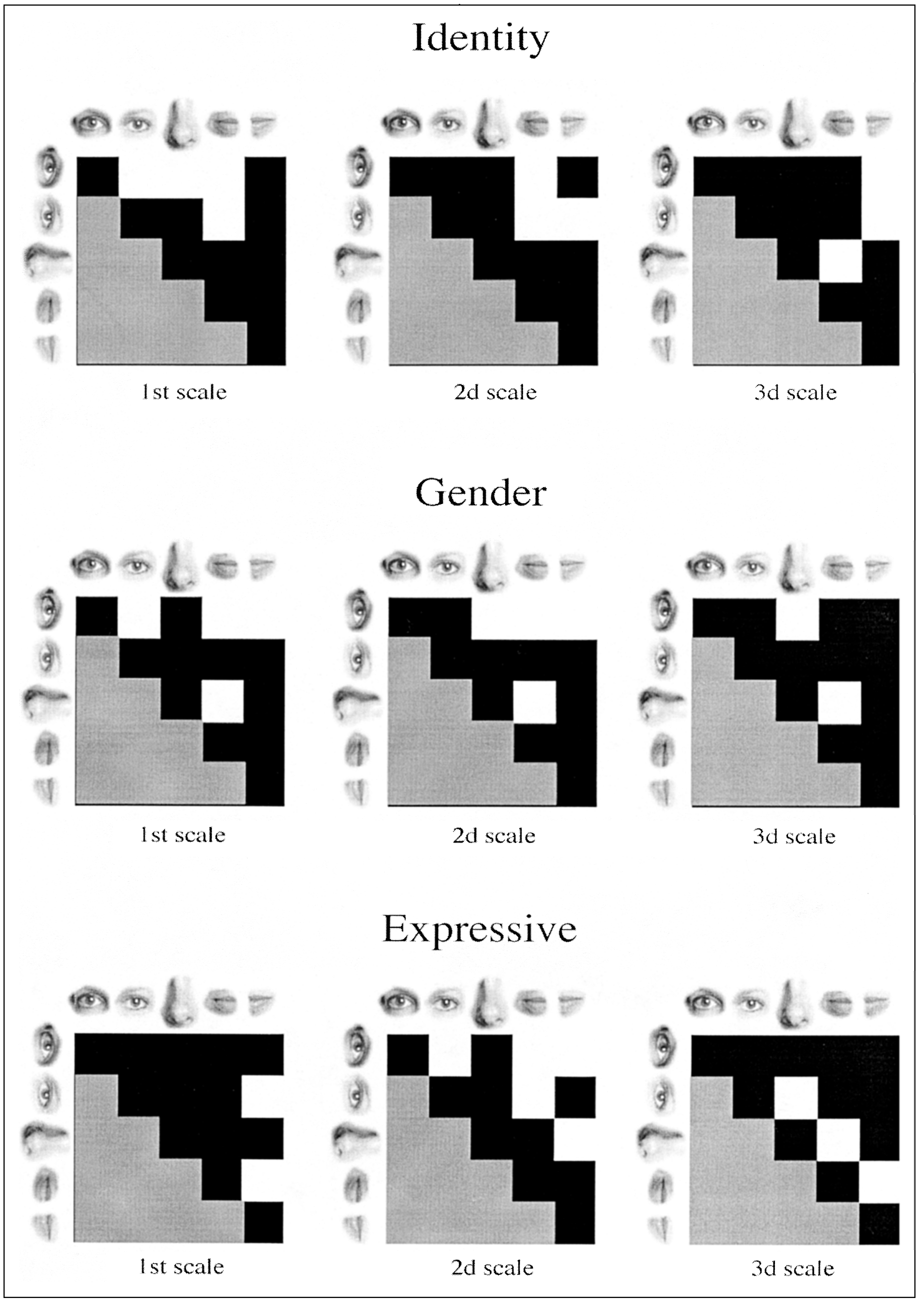

Fig. 5. Use of feature conjunctions in the identity, gender, and expressiveness tasks. At each scale, the white squares indicate significant conjunctions of features that drove recognition performance. Results for the first three scales only are shown because the other scales did not reveal any significant second-order relations. Note that the DiagnosticPlane(scale) is symmetric, but the results for each conjunction are shown only once, to improve the readability of the matrices.

REFERENCES

Ahumada, A.J., & Lovell, J. (1971). Stimulus features in signal detection. *Journal of the Acoustical Society of America*, *49*, 1751–1756.

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.

Bayer, H.M., Schwartz, O., & Pelli, D. (1998). Recognizing facial expressions efficiently [Abstract]. *Investigative Ophthalmology and Visual Science*, *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.

Brainard, D.H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*, 433–436.

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 Sciences, 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.

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

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

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

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.

Henderson, J.M., Falk, R., Minut, S., Dyer, F.C., & Mahadevan, S. (2001). Gaze control for face learning and recognition in humans and machines. In T. Shipley & P. Kellman (Eds.), *From fragments to objects: Segmentation processes in vision* (pp. 463–481). New York: Elsevier.

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

Jenkins, J., Craven, B., Bruce, V., & Akamatsu, S. (1997). *Methods for detecting social signals from the face* (Technical Report No. HIP96-39). Kyoto, Japan: Institute of Electronics, Information and Communication Engineers.

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.

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

Morrison, D., & Schyns, P.G. (2001). Usage of spatial scales for the categorization of faces, objects and scenes. *Psychological Bulletin and Review*, *8*, 454–469.

Oliva, A., & Schyns, P.G. (2000). Colored diagnostic blobs mediate scene recognition. *Cognitive Psychology*, *41*, 176–210.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

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.

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

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., Goldstone, R.L., & Thibaut, J.P. (1998). The development of features in object concepts. *Behavioral and Brain Sciences*, *21*, 1–17.

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, and Cognition*, *23*, 681–696.

Simoncelli, E.P. (1997). Image and Multi-Scale Pyramid Tools [Computer software]. New York: Author.

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

Tanaka, J., & Farah, M.J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *46A*, 225–245.

Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. In R.L. Goldstone, D.L. Medin, & P.G. Schyns (Eds.), *Perceptual learning* (pp. 83–125). San Diego, CA: Academic Press.

Tanaka, J., & Sengco, J.A. (1997). Features and their configuration in face recognition. *Memory & Cognition*, *25*, 583–592.

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.

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

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.

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

Yarbus, A.L. (1965). *Role of eye movements in the visual process.* Moscow, USSR: Nauka.

(RECEIVED 4/26/01; REVISION ACCEPTED 8/17/01)