Research Report

Transmitting and Decoding Facial Expressions

Marie L. Smith,[1] Garrison W. Cottrell,[2] Frédéric Gosselin,[3] and Philippe G. Schyns[1]

[1]Department of Psychology, University of Glasgow, Glasgow, Scotland, United Kingdom; [2]Department of Computer Science and Engineering, University of California, San Diego; and [3]Départment de Psychologie, Université de Montréal, Montreal, Quebec, Canada

ABSTRACT—*This article examines the human face as a transmitter of expression signals and the brain as a decoder of these expression signals. If the face has evolved to optimize transmission of such signals, the basic facial expressions should have minimal overlap in their information. If the brain has evolved to optimize categorization of expressions, it should be efficient with the information available from the transmitter for the task. In this article, we characterize the information underlying the recognition of the six basic facial expression signals and evaluate how efficiently each expression is decoded by the underlying brain structures.*

The ability to accurately interpret facial expressions is of primary importance for humans to socially interact with one another (Nachson, 1995). Facial expressions communicate information from which one can quickly infer the state of mind of one's peers, and adjust one's behavior accordingly. Facial expressions are typically arranged into six universally recognized basic categories (fear, happiness, sadness, disgust, anger, and surprise; Ekman & Friesen, 1975; Izard, 1971) that are similar across different backgrounds and cultures (Ekman & Friesen, 1975; Izard, 1971, 1994).

In this article, we examine the basic facial expressions computationally, as signals in a communication channel between an encoding face (the transmitter of expression signals) and a decoding brain (the categorizer of expression signals). We address three main issues: How is facial information encoded to transmit expression signals? How is information decoded to categorize facial expressions? How efficient is the decoding process?

Address correspondence to Philippe G. Schyns, Department of Psychology, 58 Hillhead St., Glasgow, Scotland, G12 0DX; e-mail: philippe@psy.gla.ac.uk.

From the standpoint of signal encoding, different facial expressions should have minimal overlap in their information: Ideal signals are encoded orthogonally to one another. To understand how the brain encodes facial signals, we relied on a model to benchmark the information transmitted in each of the six basic expressions (plus neutral), and also to quantify how these signals overlap.

As a decoder, the brain initially analyzes expression signals impinging on the retina using a number of quasilinear band-pass filters, each preferentially tuned to a spatial frequency band (De Valois & De Valois, 1991). Spatial scales are therefore good candidates as building blocks for understanding the decoding of facial expression information. We applied *Bubbles* (Gosselin & Schyns, 2001) to estimate how the brain uses spatial-scale information to decode and classify the six basic facial expressions (plus neutral), and also to quantify how the information in these expressions overlaps.

From the estimates of transmitted and decoded facial information, we measured the brain's efficiency in decoding the facial expression information that is transmitted.

EXPERIMENT

Participants

Participants were 7 male and 7 female students at Glasgow University, Scotland. All had normal or corrected-to-normal vision and were paid for their participation.

Stimuli

Stimuli were produced from 5 male and 5 female faces, each displaying the six basic facial expressions and neutral (making a total of 70 stimuli normalized for location of the eyes and the mouth). Specifically, face stimuli were posed, and the posers were tutored in producing the correct expressions for the six basic emotions according to the Facial Action Coding System (FACS; Ekman & Friesen, 1978). After the images were

captured, a certified FACS coder examined all the images and rated them using the FACS system. All 70 stimuli used met the FACS criteria for the six basic emotional expressions plus neutral. These images form part of the California Facial Expressions (CAFE) database (Dailey, Cottrell, & Reilly, 2001).[1]

Procedure

On each of 8,400 trials, observers saw information from a randomly chosen face. The image features that were presented were randomly sampled from five spatial scales, using scale-adjusted Gaussian windows. Specifically, the original face picture was decomposed into five nonoverlapping spatial frequency bandwidths of one octave each (120–60, 60–30, 30–15, 15–7.5, and 7.5–3.8 cycles/image; the remaining bandwidth served as constant background). Each bandwidth was independently sampled with a number of randomly positioned Gaussian windows adjusted at each scale to reveal 6 cycles per window. The sampled information was then recombined to produce a sparse stimulus (see Fig. 1; for details of the Bubbles procedure, see Gosselin & Schyns, 2001, and Schyns, Bonnar, & Gosselin, 2002). The advantage of sampling information across scales is that local and global face cues for face processing are presented simultaneously (see Fig. 1; see Oliva & Schyns, 1997, for discussions).

Observers categorized each sparse stimulus by pressing a labeled key on a computer keyboard. Stimuli remained on screen until response. The sampling density (i.e., the total number of Gaussian windows) was adjusted on each trial, independently for each expression, to maintain 75% correct categorization.

To benchmark the information available for performing the task, and therefore to be able to rank human use of information for each expression and scale, we built a model observer. The model was submitted to the same experiment as the human observers, using the average values derived from our human observers as parameters (i.e., accuracy for each expression, number of information samples per expression, total number of trials). However, we added to the original stimulus an adjustable density of white noise, independently for each expression, to produce the required percentage of errors for each expression (Fig. 1 illustrates the stimulus composition for the model). For each trial, the model determined the Pearson correlation between the sparse input and each of the 70 possible original images revealed with the same bubble mask. Its categorization

response was the category of the original image with the highest correlation to the sparse input (a winner-take-all scheme). Because all pixels of the 70 images were represented in memory, the model could use all information to classify the input expressions. It therefore provides a benchmark of the available facial expression information.

Results and Discussion

Following the experiment and simulation, we performed the same analysis for the human and model observers. Independently for each expression, scale, and pixel, we first computed the number of times each pixel led to a correct categorization over the number of times the pixel was presented. We then identified which of these probabilities differed significantly from the average ($p < .05$; henceforth, *diagnostic pixel probabilities*). For each expression, we computed these diagnostic pixel probabilities independently for the 7 male and 7 female observers and the 5 male and 5 female stimuli. Because there were no significant differences between genders (in either observers or expressive faces), we pooled the data.

For each expression, the diagnostic pixel probabilities circumscribe a subspace in the input space: the information that is used to classify the expression. This subspace can easily be turned into a diagnostic filter, across spatial scales and information locations, that summarizes the "information-selecting strategy" for classifying a given facial expression. We applied this filter on the original face stimuli to reveal the effective information for each facial expression.[2] Figure 2 presents these *effective faces* for the human and model observers. Using the human and model diagnostic filtering functions, we examined how the brain encodes and decodes facial expression signals.

If the human face has evolved as an efficient transmitter of facial expression signals, the filtering functions for the different expressions should generally be minimally correlated with one another, to minimize overlap of encoded signals. Note that the model used all of the information available in the 70 stimuli to categorize their expressions. We could therefore estimate how distinguishable each expression was by calculating the Pearson correlations among the model diagnostic filtering functions.[3] Table 1 reveals that the correlations were generally low ($\mu = .28$, $\sigma = .34$), with anger being quasi-orthogonal to happiness and inversely correlated with fear. We therefore conclude that the brain transmits facial expression signals with generally low

---

[1]The images can be viewed on the Web at http://www.cs.ucsd.edu/users/gary/CAFE. Although we are aware that posed expressions are made under cognitive control and might not always activate exactly the same muscle groups as spontaneous facial expressions of emotions, we feel that this does not negate the validity of FACS-coded faces as stimuli to start a scientific investigation of emotion perception (e.g., Adolphs, Tranel, Damasio, & Damasio, 1994; Adolphs et al., 1999; Blair, Morris, Frith, Perrett, & Dolan, 1999; Morris, DeGelder, Weiskrantz, & Dolan, 2001; Phillips et al., 1997; Vuilleumier, Armony, Driver, & Dolan, 2003; Winston, Vuilleumier, & Dolan, 2003). The use of posed expressions only constrains the generalization of the results.

[2]The diagnostic filter comprises five masks (one per spatial frequency bandwidth), each pixel of which can take one of two possible values: 1 if it is diagnostic, 0 if it is not. To apply the diagnostic filter to a face, we decomposed the face into five bandwidths, multiplied each bandwidth by its smoothed diagnostic mask, and recombined the partial products into an *effective face*. This operation would be illustrated in Figure 1 if the random Gaussian windows represented the diagnostic pixels.

[3]To correlate the diagnostic filtering functions, we vectorized the five bandwidths of diagnostic filters for each expression and correlated the resulting vectors with each other. These calculations produced the values in Tables 1 and 2 and Figure 2.
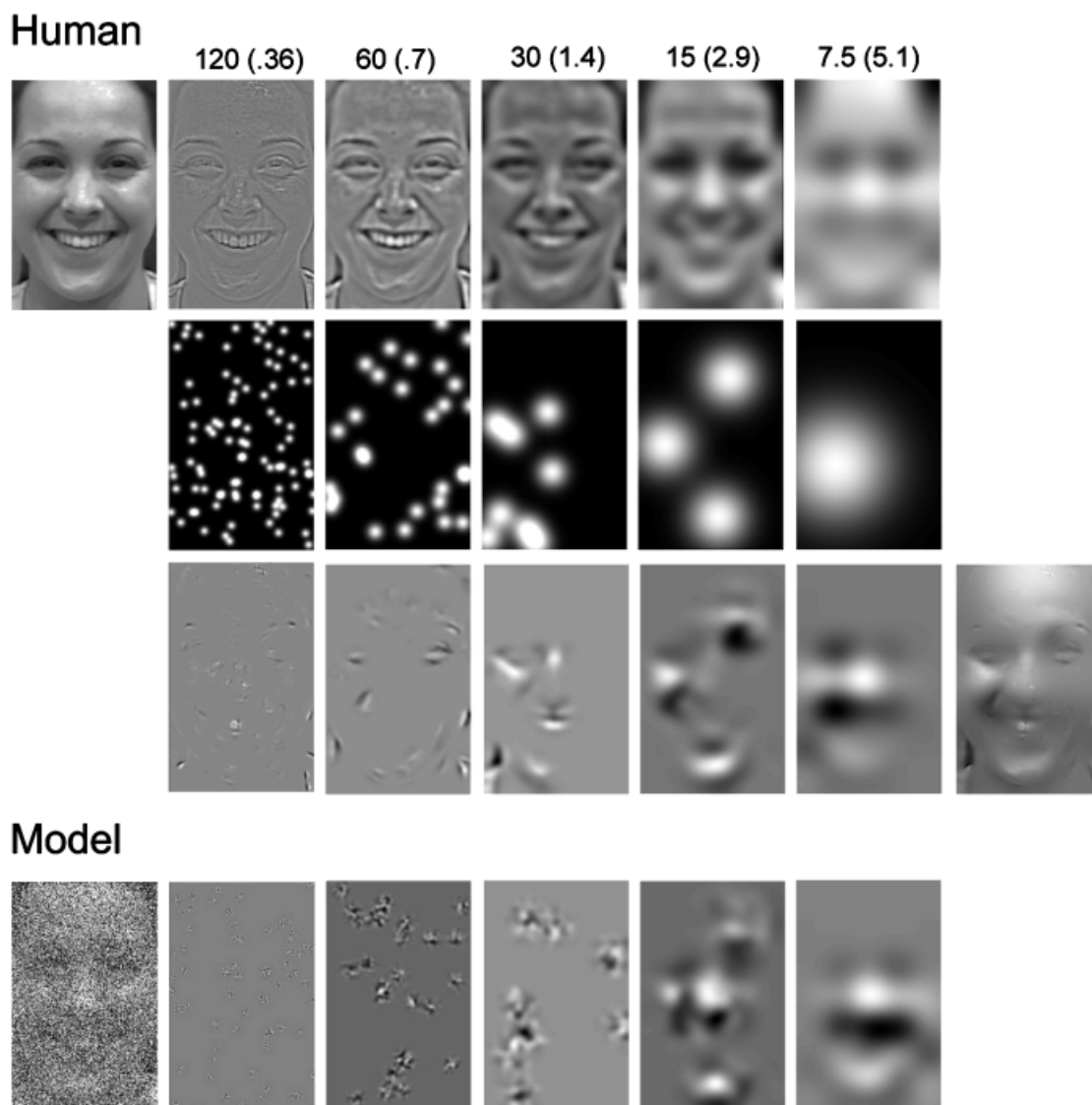
Fig. 1. Illustration of the stimulus-generation process. The upper portion of the figure shows how the bubbled stimuli were generated. First, as shown in the top row, each original face was decomposed into five spatial frequency bandwidths of one octave each (120 to 7.5 cycles/image). Each bandwidth was then independently sampled with randomly positioned Gaussian windows (0.36 to 5.1 cycles/deg of visual angle). The second row illustrates the windows in each bandwidth, and the third row shows the resulting sampling of facial information. The sum of information samples across scales (plus a constant, nonsampled, and coarsest sixth scale) produced an experimental stimulus (e.g., the right-most picture in the third row). The bottom row illustrates how the bubbled stimuli were modified to be used for the model. White noise was added to the original picture, which was then decomposed into the five spatial scales and sampled with Gaussian windows to produce one experimental stimulus.

overlap in their information. Note that the correlations themselves are based on the locations of the diagnostic information for the expressions, not on the expression information itself. Hence, our correlations values probably overestimate the correlations between the expressions. For example, although the diagnostic region is around the mouth area for both the happy and the surprise expressions, the actual information for these two expressions is quite different (open mouth with teeth in happy, empty open mouth in surprise).

On the receiving end, if the human brain has evolved to efficiently decode facial expressions, then its decoding routines should seek to minimize remaining ambiguities (technically, to orthogonalize the input classes; see Barlow, 1985, for a generic version of this point). For example, the expressions of fear and surprise are transmitted with highest overlap (.87) and are therefore ambiguous: Information from the eyes and information from the mouth are used in both transmissions (compare the human vs. model effective faces in Fig. 2). Decoding routines should seek to further decorrelate these signals to reduce their overlap and enhance categorization performance.

Interestingly, analyses of the human filtering functions revealed that they had lower correlations with each other overall
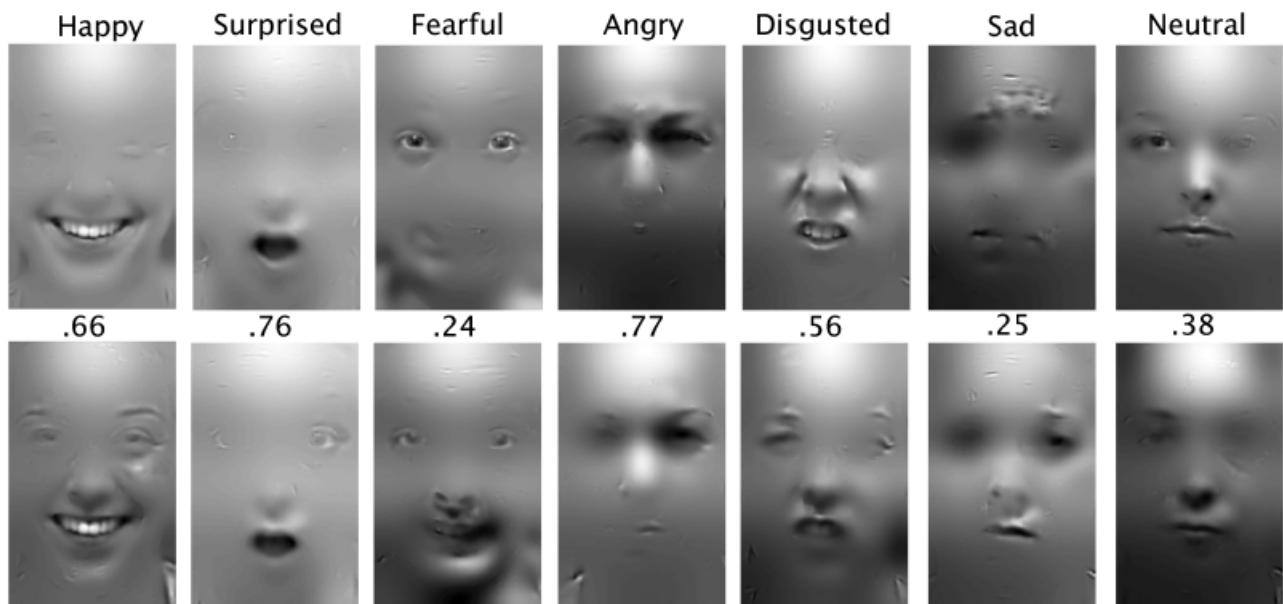
Fig. 2. Effective faces representing diagnostic filtering functions for the human (top) and model (bottom) observers. For each expression, we derived an independent diagnostic filtering function by locating, independently at each scale, the pixels leading to performance significantly ($p <$ .05) above 75% correct. We smoothed the resulting scale-specific filters and multiplied them by a sample stimulus image to produce each of the images shown here. The numbers represent the Pearson correlations between the estimated diagnostic filtering functions of the human and model observers. Higher correlations indicate higher adaptation to image information statistics. All reported correlations are between the filtering functions (not shown here), not between the applications of the filters to specific faces (i.e., the effective faces). These correlations correspond to an upper bound, and might be lower if the filters were more thoroughly characterized (e.g., with orientation).

than the model filtering functions (Table 2; $\mu = .12$, $\sigma = .25$), with anger, fear, and sadness being quasi-orthogonal to all other expressions. In the case of fear and surprise, humans use decorrelated subsets (.25) of the overlapped (.87) available information (eyes for fear and mouth for surprise; cf. the human vs. model effective faces on Fig. 2). This confirms that the decoding structures of the brain further disambiguate (i.e., orthogonalize) dissimilar inputs.

The results thus far show that the brain transmits facial expression signals with low overlap, but also that these are further decorrelated when categorized. We now turn to the relationship between transmitted and decoded facial expression signals to estimate the sensitivity of the decoder to the statistic of the transmitted information. To this end, we calculated the Pearson correlation between the human and model filtering functions for

TABLE 1
Pearson Correlations of the Model Filtering Functions

Expression	Neutral	Happy	Surprised	Fearful	Disgusted	Angry	Sad
Neutral	1	.34	.14	.12	.35	.31	.30
Happy		1	.71	.69	.71	−.09	.29
Surprised			1	.87	.54	−.24	0
Fearful				1	.66	−.36	−.08
Disgusted					1	−.12	.25
Angry						1	.57
Sad							1

each facial expression. The positive correlations ($\mu = .52$, $\sigma = .23$; see Fig. 2) indicate a sensitivity of decoding brain structures to the statistics of the information available, a desirable property in recognition tasks that have been controlled by strong evolutionary pressures. The sensitivity to information statistics can be finely measured for each facial feature with a pixel-wise comparison of the human and model filtering functions, which reflects the optimality of information use.

The optimality of information use is defined here as the logarithm[4] of a pixel-wise division of the human filtering function by that of the model. The results of this analysis are shown in Figure 3. Light blue corresponds to values close to 0, indicating optimal use of the information available to categorize a given expression (e.g., the mouth in the surprised expression). Dark blue corresponds to values below 0, indicating a suboptimal use of the available information (e.g., the left eye in the happy expression, parts of the mouth in the fearful expression). Red and yellow regions (positive values) indicate a greater use by humans than model observers of information that is not optimal for the task (a reflection of human biases; e.g., bias to the intersection of the lower forehead and eyebrows in the anger expression and bias to the region surrounding the nose in the disgust expression). Optimality can be further assessed by the spread of the distribution of optimality values around 0 (see Fig. 3).

---

[4]We used logarithms to compress the outcomes of the division—that is, to prevent small values in the denominator from causing misleadingly large values.
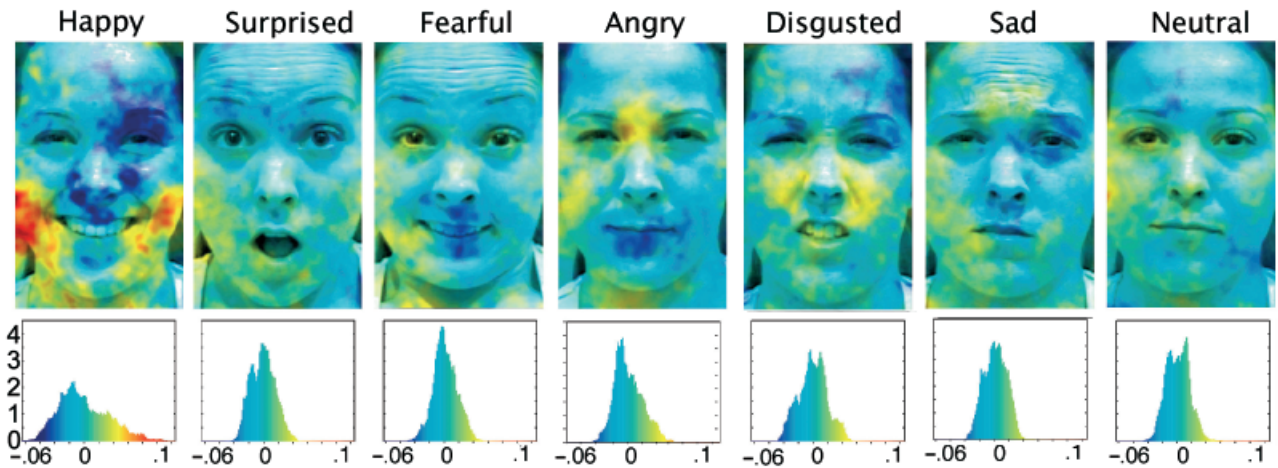
Fig. 3. Optimality of information use for each of the seven facial expressions. We define optimality as the logarithm of a pixel-wise division of the human filtering function by that of the model. The images indicate how optimally each region of the face is used for categorization, and the graphs show the distributions of the optimality values. Light blue corresponds to values close to 0, indicating optimal information use and optimal adaptation to image statistics; dark blue corresponds to negative values, indicating suboptimal information use. Red and yellow regions (positive values) indicate a greater use by humans than model observers of information that is not optimal for the task (a reflection of human biases).

The larger the spread (e.g., happy), the less optimal the use of information is. The direction of the spread toward positive or negative values indicates whether the trend is for biased or suboptimal use, respectively.

CONCLUSIONS

Our comparative analyses of the diagnostic filtering functions of the human and the model observers suggest that the face, as a transmitter, evolved to send expression signals that have low correlations with one another and that the brain, as a decoder, further decorrelates and therefore improves these signals (i.e., the effective faces in Fig. 2). Such decorrelated signals (the human filtering functions) constitute optimized inputs that can be used to isolate the specific response of specialized brain structures to the facial features transmitting facial expression signals (Blair, Morris, Frith, Perrett, & Dolan, 1999; Morris et al., 1996; Phillips et al., 1997; Vuilleumier, Armony, Driver, & Dolan, 2003; Winston, Vuilleumier, & Dolan, 2003). A direct practical implication of our results is that functional magnetic resonance imaging studies with effective faces (because they are decorrelated input signals) could in principle tease apart the brain structures (if any) that are specialized for the processing of a specific expression (see Luan Phan, Wager, Taylor, & Liberzon, 2002, for a discussion).

From a theoretical viewpoint, the idea that the face has evolved to transmit orthogonal signals raises interesting questions about how the expression of emotion signals and the development of facial muscle groups have co-evolved. In principle, the repertoire of emotion signals to be transmitted by the face has been evolutionarily constrained by the skeletal and muscular movements of the face as an encoder, by pressures such as the evolutionary advantage of decoding expressions from long viewing distances, and by the generic computational requirement of transmitting decorrelated signals. A promising research avenue could derive the categorization threshold of each facial expression in terms of viewing distance and determine the skeletal and muscular movements that are involved in these transmissions.

TABLE 2
Pearson Correlations of the Human Filtering Functions

Expression	Neutral	Happy	Surprised	Fearful	Disgusted	Angry	Sad
Neutral	1	.15	.09	.09	.43	.21	.11
Happy		1	.73	.17	.42	−.30	.19
Surprised			1	.25	.29	−.32	−.04
Fearful				1	−.19	−.02	−.06
Disgusted					1	−.05	.11
Angry						1	.27
Sad							1

REFERENCES

Adolphs, R., Tranel, D., Damasio, H., & Damasio, A. (1994). Impaired recognition of emotion in facial expressions following bilateral damage to the human amygdala. *Nature*, *372*, 669–672.

Adolphs, R., Tranel, D., Hamann, S., Young, A.W., Calder, A.J., Phelps, E.A., Anderson, A., Lee, G.P., & Damasio, A.R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. *Neuropsychologica*, *37*, 1111–1117.

Barlow, H.B. (1985). The role of single neurons in the psychology of perception. *Quarterly Journal of Experimental Psychology*, *37A*, 121–145.

Blair, R.J., Morris, J.S., Frith, C.D., Perrett, D.I., & Dolan, R.J. (1999). Dissociable neural responses to facial expressions of sadness and anger. *Brain*, *122*, 883–893.

Dailey, M., Cottrell, G.W., & Reilly, J. (2001). *California facial expressions, CAFE*. Unpublished digital images, University of California, San Diego, Computer Science and Engineering Department.

De Valois, R.L., & De Valois, K.K. (1991). *Spatial vision*. New York: Oxford University Press.

Ekman, P., & Friesen, W.V. (1975). *Unmasking the face*. Englewood Cliffs, NJ: Prentice Hall.

Ekman, P., & Friesen, W.V. (1978). *The Facial Action Coding System (FACS): A technique for the measurement of facial action*. Palo Alto, CA: Consulting Psychologists Press.

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

Izard, C. (1971). *The face of emotion*. New York: Appleton-Century-Crofts.

Izard, C.E. (1994). Innate and universal facial expressions—evidence from developmental and cross-cultural research. *Psychological Bulletin*, *115*, 288–299.

Luan Phan, K., Wager, T., Taylor, S.F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: A meta-analysis of emotion activation studies in PET and fMRI. *NeuroImage*, *16*, 331–348.

Morris, J.S., DeGelder, B., Weiskrantz, L., & Dolan, R.J. (2001). Differential extrageniculostriate and amygdala responses to presentation of emotional faces in a cortically blind field. *Brain*, *124*, 1241–1252.

Morris, J.S., Frith, C.D., Perrett, D.I., Rowland, D., Young, A.W., Calder, A.J., & Dolan, R.J. (1996). A differential neural response in the human amygdala to fearful and happy facial expressions. *Nature*, *383*, 812–815.

Nachson, I. (1995). On the modularity of face recognition—the riddle of domain specificity. *Journal of Clinical and Experimental Neuropsychology*, *17*, 256–275.

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

Phillips, M.L., Young, A.W., Senior, C., Brammer, M., Andrew, C., Calder, A.J., Bullmore, E.T., Perrett, D.I., Rowland, D., Williams, S.C.R., Gray, J.A., & David, A.S. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature*, *389*, 495–498.

Schyns, P.G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science*, *13*, 402–409.

Vuilleumier, P., Armony, J.L., Driver, J., & Dolan, R.J. (2003). Distinct spatial frequency sensitivities for processing faces and emotional expressions. *Nature Neuroscience*, *6*, 624–631.

Winston, J.S., Vuilleumier, P., & Dolan, R.J. (2003). Effects of low-spatial frequency components of fearful faces on fusiform cortex activity. *Current Biology*, *13*, 1824–1829.