Cognition xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Neural computation as a tool to differentiate perceptual from emotional processes: The case of anger superiority effect

Martial Mermillod [a,*], Nicolas Vermeulen [b], Daniel Lundqvist [c], Paula M. Niedenthal [d]

[a] Université Blaise Pascal, LAPSCO – CNRS UMR 6024, 34, Avenue Carnot, 63037 Clermont-Ferrand, France
[b] National Fund for Scientific Research, Université Catholique de Louvain at Louvain-la-Neuve, Belgium
[c] Department of Clinical Neuroscience, Karolinska Institute, Sweden
[d] National Center for Scientific Research, Université Blaise Pascal, France

A R T I C L E I N F O

Article history:
Received 30 November 2007
Revised 16 September 2008
Accepted 12 November 2008
Available online xxxx

Keywords:
Neural Computation
Emotion
Perception
Cognition
Anger superiority effect

A B S T R A C T

Research findings in social and cognitive psychology imply that it is easier to detect angry faces than happy faces in a crowd of neutral faces [Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd – An anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917–924]. This phenomenon has been held to have evolved over phylogenetic development because it was adaptive to quickly and accurately detect a potential threat in the environment. However, across recent studies, a controversy has emerged about the underlying perceptual versus emotional factors responsible for this so-called anger superiority effect [Juth, P., Lundqvist, D., Karlsson, A., & Ohman, A. (2005). Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. *Emotion, 5*(4), 379–395; Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception, 25*(9), 1091–1108]. To tease apart emotional and perceptual processes, we used neural network analyzes of human faces in two different simulations. Results show that a perceptual bias is probably acting against faster and more accurate identification of anger faces compared to happy faces at a *purely perceptual level*. We suggest that a parsimonious hypothesis related to the simple perceptual properties of the stimuli might explain these behavioral results without reference to evolutionary processes. We discuss the importance of statistical or connectionist analysis for empirical studies that seek to isolate perceptual from emotional factors, but also learned vs. innate factors in the processing of facial expression of emotion.

1. Introduction

Recently a number of theorists have proposed that the human emotion system evolved to quickly and accurately respond to signs of threat in the social environment (e.g., Hansen & Hansen, 1988). As a perceptual cue, facial expressions convey crucial information about possible social threat. This would suggest that the human emotional system is biased toward more efficient detection of angry facial expressions in the social environment (for instance, in a crowd of other faces). In an initial demonstration, Hansen and Hansen (1988) showed that experimental participants were particularly efficient at detecting an angry facial expression in a crowd of neutral faces. However, enthusiasm for this so-called "anger superiority effect" was tempered by follow up work of Purcell et al. (1996), which showed that the specific angry face used in the Hansen and Hansen studies possessed (anger-unrelated) attention-grabbing features, and that when the confound was controlled, the anger superiority effect disappeared. Since then, Öhman, Lundqvist, and Esteves (2001) reexamined the anger superiority effect using perceptually controlled schematic faces (drawn schematic faces varying only at the level of the eyebrows, eyes and mouth), and found evidence supportive of the basic phenomenon.

* Corresponding author. Tel.: +33 473406254.
E-mail address: Martial.Mermillod@univ-bpclermont.fr (M. Mermillod).

0010-0277/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2008.11.009

Because the anger superiority effect has not been demonstrated convincingly with real human faces, these differing findings have raised the possibility of a perceptual bias in favor of the detection of happy faces that competes with the faster and more accurate recognition of angry faces embedded among neutral expressions. Indeed, in a more recent study, Juth et al. (2005), using pictures of human faces (Fig. 1), found that happy faces were more quickly and accurately detected than angry faces.

Furthermore, this happiness superiority effect was reversed for schematic faces, raising once more the possibility that a natural perceptual bias (that is not present in controlled schematic faces) overrides the emotional factors involved in the anger superiority effect. In other words, drawn schematic faces do not have any perceptual variance: there is only one angry and one happy face to constitute the matrices. This is not the case for real faces; each real face, even in a set expression the same emotion, is different. Such is true, for instance, of the human faces used by Juth et al. (2005) which were nonetheless carefully controlled for perceptual factors such as color, lighting conditions, background or clothing. Thus, a perceptual bias related to the simple statistic variability of real human faces constituting the different emotional categories may exist and raises the question of the possibility of generalizing a possible anger superiority effect in the processing of real-life facial stimuli. In other words, happy faces might be statistically more differentiated than angry faces from a crowd of neutral faces and therefore detected more efficiently. This phenomenon is illustrated in Fig. 2 in order to explain precisely what do we mean by "pure perceptual factors".

In the example provided in the left part of the graph, the statistical distribution of exemplars from two categories (for instance, category A for happy and category B for neutral faces) are well-differentiated. There is a little overlap between the two categories, meaning that the average similarity between the two categories is low. In contrast, the right part of the graph illustrates two categories that are more difficult to differentiate by a statistical or connectionist network in the perceptual space provided by two

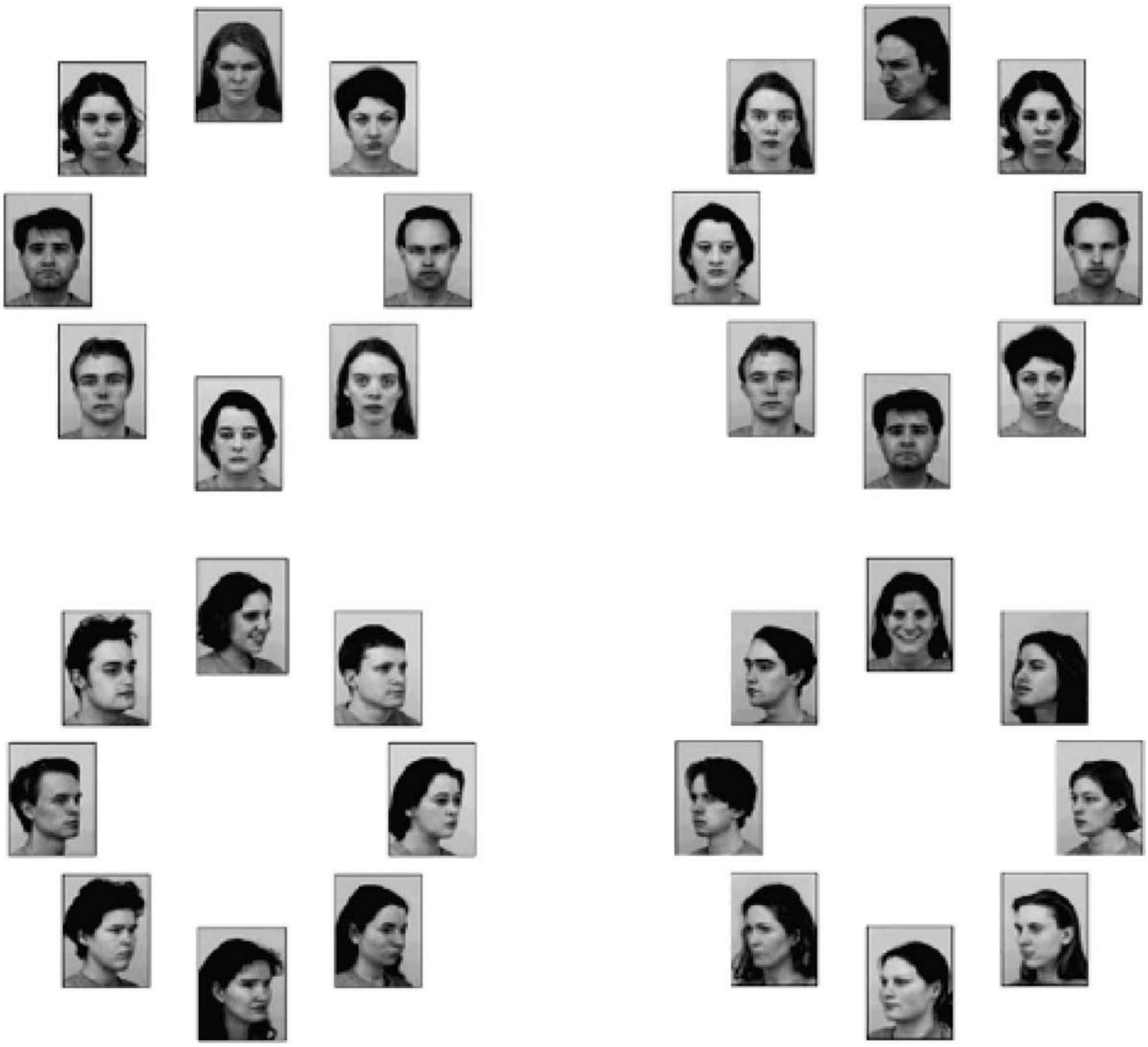

Fig. 1. Four examples of the photographic facial arrays that were used by Juth et al. (2005).

Fig. 2. Example of perceptually distinct (left part of the graph) or overlapping categories (right part of the graph). The bottom part of the figure represents our specific hypothesis in this paper (namely that happy faces are better differentiated from over angry but also neutral faces).

dimensions (for instance, category A for angry and category B for neutral faces).

The bottom part of Fig. 2 illustrates the specific hypothesis of a perceptual bias acting against superior recognition of angry faces in the "face in the crowd paradigm". We assume that happy faces are perceptually easier than angry faces to differentiate from neutral faces, resulting of a greater overlap of perceptual features between angry and neutral faces than between happy and neutral faces. Note that this figure is a simple example of a visual representation of three categories in a two dimensional space. In the current paper, we work in the 56-dimensional space of the perceptual layer produced by Gabor receptive fields. Such high-dimensional space is more difficult to represent (Pothos & Close, 2008) but the basic idea is exactly the same as the example represented in Fig. 2, generalized to a greater than two-dimensional space.

Concerning cognitive and emotional factors, the goal of an hypothetic fear system aimed at distinguishing rapidly angry from neutral faces is *to find efficiently non-linear boundaries* (or even a linear boundaries if these are sufficient to resolve the task) across the three categories. Thus, the basic assumption of the anger superiority effect is that the human cognitive system has evolved to more efficiently recognize anger than happy faces among a crowd of neutral distracters because of survival purpose, irrespective to the perceptual structure of the three categories. However, we suggest that a more parsimonious explanation, related to the simple perceptual distribution of the inputs, is able to account for the previous behavioral data.

In order to test the statistical properties of the inputs, we needed a computational tool. In addition to statistic analysis, neural computation is a very useful method for the present purpose because it is *not submitted to possible* *innate influence*. Different perceptual data can be submitted to exactly the same artificial cognitive system to test the hypothesis that the perceptual variability of the three categories is sufficient to simulate the previous behavioral results. We tested the statistical properties of these human facial expressions at a purely perceptual level using the same stimuli (KDEF database, Lundqvist & Litton, 1998) and the same experimental design (distractors and target faces directed or averted toward the observer) as those used in the extant major publications.

It is important to note that we do not argue that the present neural computational modeling represents a neurobiological model of happy or angry face perception per se. Rather, we recognize it as a computational tool that allows us to evaluate the distinctive statistical properties of facial expressions of emotion. In our work we used one of the most broadly accepted models of vision, based on Gabor wavelet filtering, applied to face perception (Lyons, Akamatsu, Kamachi, & Gyoba, 1998; Lyons, Budynek, & Akamatsu, 1999; Wiskott, 1997). We associated the vision model with one of the most frequently used models of cognitive processes, based on connectionist neural networks, in order to stay as close as possible to the main references in neural network modeling applied to recognition of facial expressions of emotion (Dailey & Cottrell, 1999; Dailey, Cottrell, Padgett, & Adolphs, 2002; Lyons et al., 1999; Zhang, Lyons, Schuster, & Akamatsu, 1998).

It should be pointed out that the results of such modeling are determined largely by the perceptual structure of the stimuli rather than by the algorithm used. For example, the data in the right panel of Fig. 2 will be more difficult to categorize than the data reported on the left part of the figure when submitted to a back-propagation algorithm (i.e., multi-layer perceptron), but also to other neural network algorithms such as self-organizing map (SOM), radial basis function (RBF) networks, or a single layer auto-encoder based on the Hebbian learning rule (Abdi, Valentin, Edelman, & O'Toole, 1995). Even statistical analysis such as discriminant analysis (Lyons et al., 1999), multi-dimensional scaling or PCA will provide similar results depending on the statistic variability of the data. We chose to use a standard back-propagation algorithm for three reasons. First, the introduction of parallel distributed processes (PDP) by Rumelhart, Hinton, and McClelland (1986) constituted a scientific revolution in the understanding of cognitive processes. PDP models provide an innovative understanding of a wide range of psychological data (McClelland & Rogers, 2003). Second, at a computational level, PDP networks constitute a powerful training algorithm that can reveal non-linear boundaries even in complex high-dimensional spaces (Rumelhart et al., 1986). Third, back-propagation algorithms are one of the most widely used and standardized type of neural network in the field of cognitive science (McClelland & Rogers, 2003). More importantly, with respect to the general aim of the current paper, previous work has shown that this technique is particularly efficient in the recognition and categorization of emotional facial expressions (Dailey et al., 2002; Lyons et al., 1999; Zhang et al., 1998).

Concerning the choice of parameters, again, an infinite number of parameters are available. For example, changing the learning rate or the momentum will change the speed or the efficiency of training. Therefore, we chose the same parameters as other publications (French, Mareschal, Mermillod, & Quinn, 2004; Mermillod, Chauvin, & Guyader, 2004; Mermillod, Guyader, & Chauvin, 2005a, 2005b) with the general goal of having the maximum of efficiency of the neural network for visual categorization tasks. However, it is important to note that changing a parameter will change the quality of training *in the same manner* across the different experimental conditions (actually, the back-propagation algorithm is highly standardized and there is actually less possibility of misrepresenting the data with such an algorithm than with ANOVA for statistical analyzes for example). In other words, increasing the learning rate increases the speed of training equally in the different conditions. Nonetheless, overlapping categories will always be more difficult to recognize than well-differentiated categories (see Fig. 2) as long as the same parameters are *kept constant across all training conditions*.

The first simulation presented below is a general test of the neural network's classification of the happy, anger and neutral faces used by Juth et al. (2005), when the three categories are presented together to the network. Simulation 2 is a specific simulation of the experimental design used by Juth et al. (2005), which consists of detecting a discrepant face (either anger or happy face) in a crowd of neutral distractors.

2. Simulation 1

2.1. Method

2.1.1. Connectionist network

The connectionist model was decomposed in two main stages. The first component was a perceptual model of vision simulating V1 neuron receptive fields that are sensitive to different orientations and spatial frequency channels. It has been shown that visual information can be efficiently compressed by Gabor filter decomposition and, most importantly, with a remarkable biological plausibility (Jones & Palmer, 1987; Jones, Stepnoski, & Palmer, 1987). Dailey et al. (2002), Lyons et al. (1999), Zhang et al. (1998) all demonstrated that Gabor filters combined with back-propagation neural networks or discriminant analyzes are able to successfully classify human facial expressions in a reliable way comparable to human participants.

Our current neural network is very similar to their model except that Gabor filters were applied in the frequency domain (or Fourier domain) instead of the spatial domain. Multiplying a Gabor filter in the Fourier domain is equivalent to convolving it into the spatial domain. One characteristic of this method is that it does not coding for spatial location: Applying Gabor filters in the spectral domain renders the representation of the image (i.e., the output computed by the Gabor filters) translation invariant, exactly like V1 complex cells which are pooling information of different V1 simple cells sensitive to the same scale and orientation but at different place of the visual field (De Valois & De Valois, 1988). The advantage of this method is that it avoids the use of PCA to compress visual information. In other words, the compression step necessary for subsequent neural computation algorithms occurs at the level of the Fourier transform instead of PCA. Note that the *dimensionality reduction* of the input space realized by PCA (Dailey et al., 2002; Lyons et al., 1999) is also losing spatial location. Thus, the Fourier transform avoids the use of PCA to reduce the large amount of information when Gabor filters are applied on a sliding window in the spatial domain.

We applied a single bank of fifty-six Gabor filters corresponding to seven spatial frequency bands (one octave per spatial frequency channel) and eight different orientations ($0, \pi/8, 2\pi/8, 3\pi/8, 4\pi/8, 5\pi/8, 6\pi/8, 7\pi/8$), with respect to biological data (De Valois & De Valois, 1988). The energy coefficients provided by the Gabor filters were computed by multiplying the local energy spectra by the function of the Gabor filter and taking the average energy value provided by the filter (computational details are provided in the Appendix). Thereafter, the second component is a back-propagation neural network whose aim is to classify the output vectors provided by the Gabor filters (Dailey & Cottrell, 1999; Dailey et al., 2002; Mermillod, Vuilleumier, Peyrin, Alleysson, & Marendaz, in press; Mermillod et al., 2005a, 2005b).

The neural network was used for Simulation 1 in a hetero-associative mode similar to Dailey et al. (2002) as described in the Section 2.1.3. The synaptic weights were adjusted by means of the standard back-propagation algorithm (Rumelhart et al., 1986, see Appendix for details). As mentioned above, our connectionist network was not used here as a model of the entire human cognitive system, but rather constituted a computational tool that allowed us to analyze the subtle and distinctive statistical properties of the facial expressions. Put differently, we used neural network modeling as a simple non-linear classification system applied to the different emotion cues provided by the facial expressions of emotion.

2.1.2. Stimuli

For all simulations, the stimuli were the original images used in Juth et al. (2005), taken from the Karolinska directed emotional faces set (KDEF, Lundqvist & Litton, 1998). These included 540 human faces (half were male and half were female faces) from three categories (60 neutral faces, 60 angry faces and 60 happy faces). Each of 60 different individuals displayed the three emotional expressions (neutral vs. angry vs. fearful expression), in a directed (full-frontal) or averted (half-profile) to $-45°$ or averted to $+45°$ viewpoint. Color images were transform to 256 gray-level scale for computational reasons and a Hann window was applied in order to avoid over-representation of cardinal orientations (due to image edges) in the spectral domain.

2.1.3. Procedure

In these first simulations, we used the standard hetero-association training algorithm to associate each of the different category exemplars with a specific output vector coding for them. All training categories were learned together by the neural network. For each stimulus, 56 Gabor filters were applied on the original image in order to compute 56 descriptors coding for the local energy spectra of the image. Then, the length-56 energy vectors were associated by a 3-layer back-propagation network with their suitable code category (1 0 0 for angry faces, 0 1 0 for happy faces and 0 0 1 for neutral faces), and a new image from the training set was coded and associated by the neural network in an iterative process.

The network architecture consisted of 56 input units, 28 hidden units and 3 output units. Note that the size of the hidden layer, as the number of epochs or the value of the learning rate, does not have implications at a qualitative level. The only qualitative constraint concerning the hidden layer is to produce a bottleneck from the input to the output layer in order to reduce the dimensionality from the perceptual input to the categorical output layer. At a quantitative level, the larger the hidden layer or the number of epochs, the better the training of the neural network. However, any slight improvement of performance related to these basic parameters was equivalent in all training conditions. Here, while keeping exactly the same architectures and parameters for neural networks in all training conditions, we were able to apply the most standard type of connectionist network, with the most common training regime, in order to test for basic statistical properties of happy, angry and neutral stimuli. The learning rate was fixed to 0.1 and momentum to 0.9. The goal of the back-propagation network was to create non-linear category borders between the three perceptual categories.

2.1.4. Training phase

Each run began with a random selection of 90 training exemplars (30 angry, 30 happy and 30 neutral faces). Then the training consisted of associating each of the 90 exemplars with the appropriate code category for 500 epochs.

2.1.5. Test phase

After the neural network was trained simultaneously on both expression categories, it was then tested on the 30 remaining novel exemplars from the angry, happy and neutral categories. An output vector was computed by the artificial neural network after exposure to each input vector. Then, we applied a winner-take-all procedure on the output vectors. Our dependent measure was the correct classification rate produced by the neural network. Results were averaged over 50 runs of the above training-test procedure.

3. Results

3.1. Directed (full-frontal) faces

As shown in the confusion matrix presented in Table 1, the connectionist network produced an average correct categorization rate of 94% ($SE = 0.0074$) for new exemplars from the anger category, 97.1% ($SE = 0.0041$) for new exemplars from the happy category and 91.9% ($SE = 0.0084$) for new exemplars from the neutral category. There was a main effect of category of face on categorization level ($F(2.98) = 15.46$, $\text{MSE} = 0.0023$, $p < 0.001$). For the corresponding first experiment of the original behavioral study (Juth et al., 2005), the authors obtained an average correct categorization rate of 92% for happy faces and 88% for angry faces, the difference was significant. For the rest of

Table 1
Confusion matrix as a function of the input exemplar, the observed output and the direction of the processed face.

Input exemplar	Observed output		
	Anger	Happy	Neutral
	Directed (full-frontal) faces		
Anger	94.0	4.7 (.0041**)	1.3 (.0698)
Happy	2.5 (.0041**)	97.1	0.4 (.0001**)
Neutral	4.9 (.0698)	3.3 (.0001**)	91.9
	Averted (half-profile) faces to +45°		
Anger	96.3	2.6 (0981)	1.1 (.1452)
Happy	0.3 (.0981)	97.7	2.0 (.0005**)
Neutral	2.5 (.1452)	2.6 (.0005**)	94.9
	Averted (half-profile) faces to −45°		
Anger	96.8	2.3 (.004**)	0.9 (.303)
Happy	2.1 (.004**)	94.1	3.7 (.170)
Neutral	0.1 (.303)	4.3 (.170)	95.6

Note: p-Values for exhaustive HSD Tukey comparisons between each pair of correct categorization are presented parenthetically.

statistical analyzes, we have chosen to test all possible comparisons between each pair of training conditions. Because of the lost of degrees of freedom, we applied a post-hoc HSD Tukey test in order to ensure correct statistical p values. Exhaustive comparisons between each pair of training conditions for correct categorization rate are also presented in Table 1. These reveal that the difference between correct categorization rates for happy faces and angry faces was significant as well as the difference between happy faces and neutral faces, meaning that categorizing happy faces is easier than categorizing neutral or angry faces. The difference between angry and neutral faces was smaller and not significant, highlighting the fact that angry faces were harder than happy faces to differentiate from neutral faces.

3.2. Averted (half-profile) faces

Table 1 reports the results for faces averted to +45°. As for directed faces, results showed a correct categorization rate of 96.3% (SE = 0.0044) for new exemplars from the anger category, 97.7% (SE = 0.0036) for new exemplars from the happy category and 94.9% (SE = 0.0059) for new exemplars from the neutral category. As for directed faces the main effect of category faces on correct categorization levels was significant ($F(2.98) = 7.9$, MSE = 0.001, $p < 0.001$). Table 1 shows that happy faces were significantly easier to categorize than angry and neutral faces.

Finally, Table 1 shows the results for faces averted to −45°. The connectionist network produced an average correct categorization rate of 96.8% (SE = 0.006) for new exemplars from the anger category, 94.1% (SE = 0.006) for new exemplars from the happy category and 95.6% (SE = 0.0047) for new exemplars from the neutral category. There was a main effect of category of face on categorization levels ($F(2.98) = 5.46$, MSE = 0.002, $p < 0.01$). However, for averted −45° faces, Table 1 reveals that happy faces were harder to categorize from angry faces but also that neither happy nor angry faces produced a significant difference from neutral faces.

4. Discussion

Results generated by the standard hetero-association process suggest that the categorization of happy faces by an artificial system based on parallel and distributed processes should be easier, at a purely perceptual level, than categorizing angry or neutral faces. This pattern was obtained for both directed and +45° averted faces. For −45°, the categorization of happy and angry faces was not superior to neutral faces. Therefore, the perceptual bias toward better recognition of happy faces was shown as particularly acute for directed and +45° averted faces.

However, these results alone are not conclusive. In their behavioral findings, Juth et al. (2005) showed that happy faces were easier and more accurately recognized than angry faces *in a crowd of neutral faces*. In the previous simulations, we tested the performance of a non-linear model of categorization among the three categories together, but not from one particular (happy or angry face) to another particular category (neutral face). The next series of simulations was performed to provide a (better) fit of the original experimental conditions.

5. Simulation 2

5.1. Method

5.1.1. Connectionist network and stimuli

We used an original training algorithm based on an auto-encoder (the learning rule of the back-propagation remains the same). This training algorithm was previously used in other research (French et al., 2004; Mareschal, French, & Quinn, 2000) to simulate infants' perceptual categorization abilities (i.e., without semantic knowledge). The training algorithm operates as follows. First, a stimulus (i.e., an energy vector produced by the Gabor filters) is presented to the input layer of the network; the resulting activation is propagated through the neural network and produces an output activation; then, this output activation is associated with the theoretically correct vector (the same as the input vector) and error is back-propagated to reduce the discrepancy between the observed output vector and the theoretical output expected; and then a new vector from the same category is auto-associated with itself, and so forth. The aim of the algorithm is to create at the hidden layer level an "internal representation" of a specific category. In the following simulations, this internal representation, as in the original Juth et al. (2005) paper is based on the neutral category faces.

Finally, we exposed the trained neural network to new exemplars, from either the same (neutral) or other stimulus categories (angry and happy faces), and we measured the Euclidean distance between the expected output and the observed output. The basic aim of this simulation is to determine the distance between happy vs. angry faces to the internal representation of neutral faces. The higher the Euclidean distance between novel category exemplars, the higher the perceptual discrepancies between the two categories (French et al., 2004). Except for this training algorithm, the neural network learning rule and stimuli were identical to Simulation 1.

5.1.2. Procedure

In contrast to the procedure of Simulation 1, here the length-56 energy vectors were not associated with arbitrary code categories but rather with themselves. Therefore the network architecture consisted of 56 input units, 52 hidden units and 56 output units. The learning rate was fixed to 0.1 and momentum to 0.9.

5.1.3. Training phase

Each run began with a random selection of 30 training exemplars from neutral category faces. Then the training consisted of associating each of these 30 training exemplars with themselves for 500 epochs.

5.1.4. Test phase

After the neural network was trained on neutral category expressions, it was then tested on the 30 remaining exemplars from the novel neutral, angry and happy categories. We computed the Euclidean distance between the observed output produced by the neural network and the expected output (i.e., that was identical to the input vector introduced into the network). Results were averaged over 50 runs of the above training-test procedure.

6. Results

6.1. Directed (full-frontal) distractor faces/directed (full-frontal) target face

As shown in Fig. 3, when target faces were presented directed toward the observer among directed distractor faces, happy faces produced an average Euclidean distance of 0.92 (SE = 0.0062) that was significantly higher (Table 2) than the Euclidean distance produced by angry faces (Euclidean distance: 0.86, SE = 0.006) and new neutral faces (Euclidean distance: 0.73, SE = 0.0046). Tukey comparisons were showing that angry faces were well-differentiated from new neutral faces, but also that happy faces were significantly more distant from new neutral faces than angry faces. This means that the perceptual distance separating

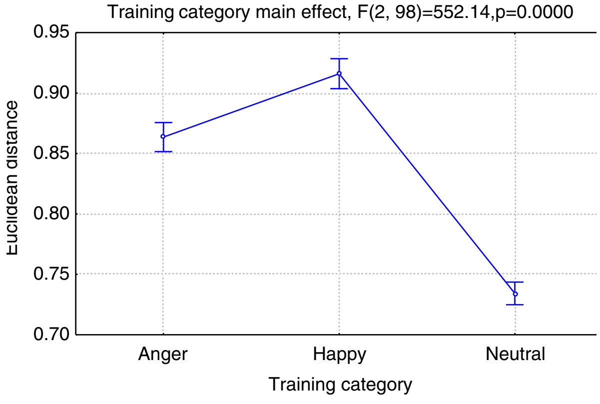

Fig. 3. Average Euclidean distance produced by directed (full-frontal) target faces after training on directed (full-frontal) distractors.

Table 2
p-Values for exhaustive HSD Tukey comparisons between each pair of Euclidean distances for directed (full-frontal) distractors/directed (full-frontal) targets.

Training category	Training category		
	Anger	Happy	Neutral
	Directed distractors/directed targets		
Anger		0.0001**	0.0001**
Happy	0.0001**		0.0001**
Neutral	0.0001**	0.0001**	

happy faces from the perceptual representation of neutral faces created at the hidden layer level of the neural network is higher than the perceptual distance separating angry faces from the perceptual representation of neutral faces. In other words, the neural network recognized happy faces as more different from neutral faces than angry faces.

6.2. Averted (half-profile) distractor faces/averted (half-profile) target face

In this simulation, distractors and target faces were presented averted to −45° or +45°. For −45° averted distractors and targets (Fig. 4), happy faces produced an average Euclidean distance of 1.10 (SE = 0.01) as compared to 1.05 (SE = 0.011) for angry faces and 0.91 for new neutral faces (SE = 0.009). As for directed faces, all differences were significant (Table 3). Similarly, for +45° averted distractors and targets, happy faces produced an average Euclidean distance of 1.13 (SE = 0.017). Once again, this value was significantly higher (Table 3) than the Euclidean distance for angry faces 1.06 (SE = 0.014) and new neutral faces (M = 1.01, SE = 0.015).

6.3. Directed (full-frontal) distractor faces/averted (half-profile) target face

In this simulation, faces were presented averted to −45° or +45° toward the observer among directed distractor faces. For −45° averted target faces, the average Euclidean distance was 1.35 (SE = 0.02) for happy faces, 1.34 (SE = 0.018) for angry faces and 1.09 (SE = 0.011) for neutral faces. The difference between happy and angry was not significant (Table 3), indicating that happy faces were as well recognized as angry faces among neutral distractors. For +45° averted target faces, happy faces produced an average Euclidean distance of 1.30 (SE = 0.015), 1.35 (SE = 0.017) for angry faces and 1.09 (SE = 0.011) for neutral faces. This difference was significant (Table 3). As shown in Fig. 5, this specific experimental condition did not replicate a better perceptual recognition of averted happy faces among a crowd of neutral averted faces. However, when +45° averted target faces and −45° averted target faces were averaged, as was the case in the original experiment, there was no remaining significant difference between angry and happy faces (as shown in synthesis results below mentioned), suggesting that the perceptual bias toward better recognition of happy faces is probably weaker or absent under this specific condition. This result partially corroborates the original results since this condi-

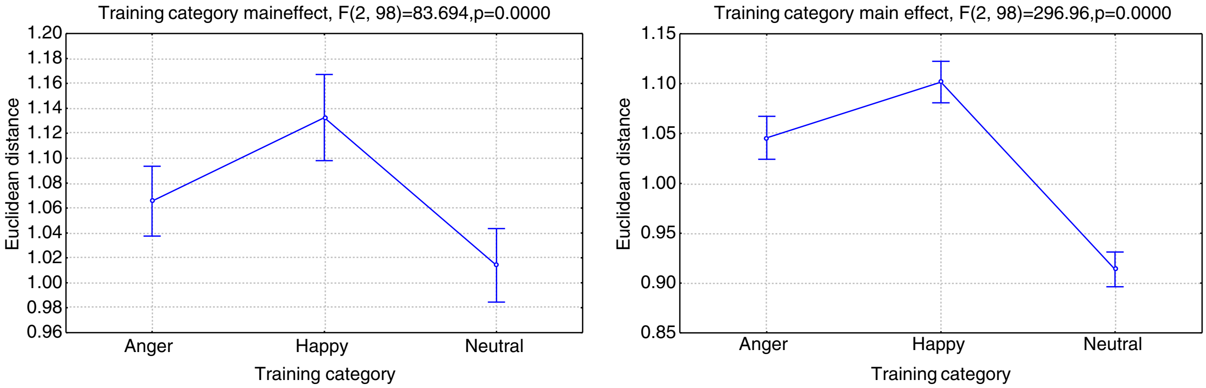

Fig. 4. Average Euclidean distances produced by averted (half-profile) target faces after training on averted (half-profile) distractors (left graph: −45°, right graph: +45°).

Table 3
p-Values for exhaustive HSD Tukey comparisons between each pair of Euclidean distances as a function of the targets and of the distractors orientations.

Training category	Test categories					
	Anger	Happy	Neutral	Anger	Happy	Neutral
	Averted −45° distractors/averted −45° targets			*Averted +45° distractors/averted +45° targets*		
Anger		0.00010**	0.00011**		0.00010**	0.00011**
Happy	0.00010**		0.00010**	0.00010**		0.00010**
Neutral	0.00011**	0.00010**		0.00011**	0.00010**	
	Directed distractors/averted −45° targets			*Directed distractors/averted +45° targets*		
Anger		0.551574	0.000105**		0.000162**	0.000105**
Happy	0.551574		0.000105**	0.000162**		0.000105**
Neutral	0.000105**	0.000105**		0.000105**	0.000105**	
	Averted −45° distractors/directed targets			*Averted +45° distractors/directed targets*		
Anger		0.006306**	0.000105**		0.000105**	0.000105**
Happy	0.006306**		0.000105**	0.000105**		0.000105**
Neutral	0.000105**	0.000105**		0.000105**	0.000105**	

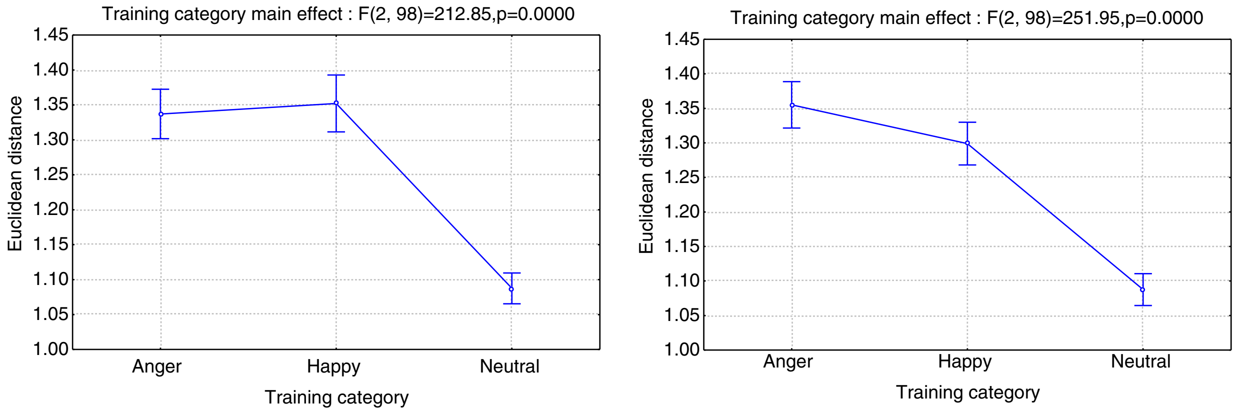

Fig. 5. Average Euclidean distances produced by averted (half-profile) target faces (left graph: −45°, right graph: +45°) after training on directed (full-frontal) distractors.

tion was the only condition where the authors observed an effect of happy faces at the level of accuracy but not at the level of reaction time (Fig. 7). Moreover, this experimental condition has no theoretical importance in the original paper (Juth et al., 2005) because they did not assume averted target faces to produce an effect among neutral directed distractors (Fig. 6).

6.4. Averted (half-profile) distractor faces/directed (full-frontal) target face

Contrary to the previous simulation, target faces were presented directed to the observer whereas distractor faces were averted to −45° or +45°. For −45° averted distractors, happy faces produced an average Euclidean distance of

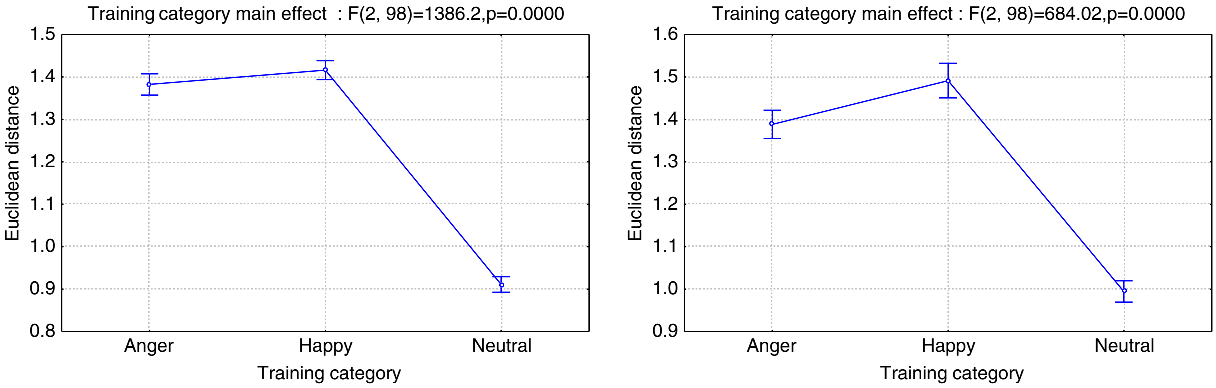

Fig. 6. Average Euclidean distances produced by directed (full-frontal) target faces after training on averted (half-profile) distractors (left graph: −45°, right graph: +45°).

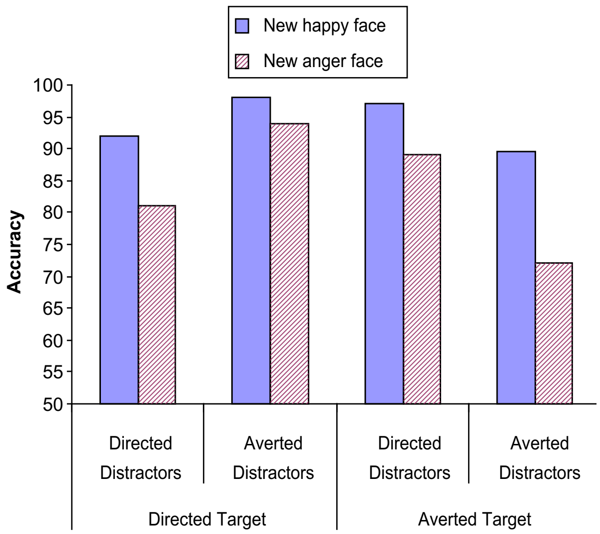

Fig. 7. Behavioral data reported by Juth et al. (2005).

1.42 ($SE = 0.011$) that was significantly higher (Table 3) than the Euclidean distance of 1.38 ($SE = 0.012$) for angry faces and 0.91 ($SE = 0.009$) for neutral faces. Similarly, for +45° averted distractors, happy faces produced an average Euclidean distance of 1.49 ($SE = 0.02$). This value was higher than the Euclidean distance for angry faces: 1.39 ($SE = 0.017$) but also compared to the Euclidean distance for neutral faces: 0.99 ($SE = 0.012$). Both differences were significant. The bias toward happy faces was therefore particularly important for these experimental conditions because the anger superiority effect was assumed to be very acute in the behavioral studies for directed faces in a crowd of averted faces. Note that, as in the original Juth et al. (2005) study, the neural network had an overall better generalization rate (i.e., lower Euclidean distance) when directions of target and distractors where congruent (directed distractors/directed target or averted distractors/averted target) compared with incongruent directions (directed distractors/averted target or averted distractors/directed target).

6.5. Synthesis and comparisons with behavioral results

The aim of this part is a synthesis of the above mentioned connectionist data and a comparison to the behavioral data (Fig. 7) obtained by Juth et al. (2005). Results in Fig. 8 show a significant perceptual bias ($p < 0.001$ with a post-hoc Tukey) for happy faces compared to angry faces in each experimental condition, except for directed distractors/averted target. The two most important conditions at a behavioral level to obtain the anger superiority effect in the “face in the crowd” paradigm were for directed target, among directed or averted distractors. Under these specific conditions we observed a clear perceptual bias toward happy faces. Connectionist simulations revealed no significant perceptual bias for anger or happy faces in directed distractors/averted target but the bias is significant for

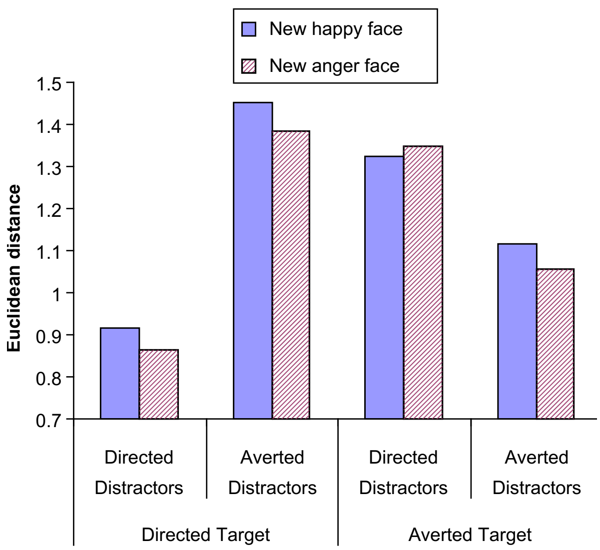

Fig. 8. Synthesis of connectionist results for each behavioral condition.

averted distractors/averted target, even if the authors do not assume anger superiority effect for averted targets.

7. General discussion

In this work, we examined the perceptual factors at play in the efficiency of detecting a face, angry as opposed to happy, in a crowd of neutral faces. Employing the stimuli used by Juth et al. (2005), but used also in the major extant papers on this topic, our first simulation showed that happy faces were more easily categorized by the network than were angry and neutral faces. This was found for directed and +45° averted faces when all categories were learned together, whereas the difference was not significant for −45° averted faces. In our second simulation, which corresponds to a more precise replication of the Juth et al. (2005) experimental design, we observed a perceptual bias for happy faces in the two most important conditions (when target faces were directed toward the observer). This could easily explain the unexpected better and faster detection of happy faces in the Juth and colleagues' studies. Note that the artificial neural network was able to generalize training from averted faces to directed faces for the different emotional categories which was an important result per se. This result is consistent with previous behavioral and computational data (Valentin & Abdi, 1996) showing that even a single layer neural network performing face recognition is able to generalize training from full-frontal to half-profile viewpoint as well as human observers. However, these results were based on orientation and identity but did not combine orientation and EFE. Further behavioral and computational data must be collected in order to explore the generalization capabilities of EFE across different viewpoint in human and artificial systems.

Our results clearly show that, even if the facial expressions used by Juth et al. (2005) were carefully controlled for color, lighting conditions or other contextual differences such as background or clothing, there exists a perceptual bias intrinsic to the expressions of happy, angry and neutral faces. This bias could constitute a parsimonious explanation of their finding of a happy superiority effect since participants could discriminate happy face more easily than angry face. The critical issue of whether the face-in-the-crowd effect should be more appropriately labeled perceptual or emotional, raised by the pioneering study nearly 20 years ago, is therefore still a topical question for behavioral studies (Hansen & Hansen, 1988, p. 923).

More generally, happy EFE seem to be easier to recognize than other EFE for human observers (Russell, 1994 for a review across different cultures). The problem with human participants was that it was very difficult to determine if this better recognition rate is produced by perceptual factors (relevant features related to happiness are more differentiated than perceptual features related to other EFE) or by more complex cognitive or emotional factors. For example, happiness is the only clear positive valence emotion across the six basic EFE. Thus, one might assume that it will be easier to recognize happy expressions if the valence has a major impact on determining the recognition of specific EFE. Another possible explanation for fast recognition of happy EFE could be that it could be very useful for humans to recognize happy expressions very rapidly, for social purpose for example. Thus our cognitive or emotional system may allocate specific resources to this task and fast detection might not be related to perceptual features but to cognitive or emotional resources. Compared to human observers, the use of neural networks (as well as statistical analysis of the stimuli) allows investigating the perceptual structure of stimuli without any references to hypothetic innate emotional processes driven by the phylogenetic development.

Contrasting with the conflicting results observed with real faces, schematic faces gave rise to a reliable anger superiority effect (Juth et al., 2005; Öhman et al., 2001). Because schematic faces are created by using identical physical features, there exist the same physical differences from neutral to angry faces and from neutral to happy faces. The problem with schematic faces is that there is no perceptual variability in the data. The three categories are represented by three single points in the perceptual space, which is completely unrealistic compared to real human faces. Thus, along with a superior perceptual control, schematic faces could always be seen as lacking ecological validity.

The generalizability of the anger superiority effect in real faces could be clearly anticipated by using perceptually controlled faces. From this perspective, neural networks could be used as a computational tool in order to select, on a perceptual basis, stimuli from different categories (e.g., emotion) but also for different stimuli (e.g., faces or scenes). This should be done in diverse paradigms such as visual search paradigms like the face-in-the-crowd. Selecting happy, angry or neutral facial expressions that are perceptually well-differentiated (see Fig. 2) is expected as a way to avoid those perceptual biases. In other words, among the different faces used in the experiment, we are able to determine by means of neural computation which of them do not overlap with other category members. Therefore, perceptual biases could be controlled on the basis of the Euclidean distances separating each category. We could, therefore, select real expressions from different emotional categories (i.e., anger, joy) that are physically (i.e., perceptually) equally distant from the neutral faces in order to test emotional processes while controlling perceptual factors.

To conclude, the present simulations provide computational evidence in support of the suggestion by Juth et al. (2005) that the happy superiority effect is perceptually driven. Neural network modeling was used as a computational tool allowing us to test the basic statistical properties of the stimuli (in determining which category is perceptually more distant from one or the other category). Results provided evidence for better and easier recognition of happy faces, compared to angry faces, among a crowd of neutral distractors (particularly when target faces are directed toward the observer). This raises major concerns about the possibility of better recognition of angry among neutral faces, *for real-life stimuli*, evolved over phylogenetic development. The recent behavioral results seem to be best explained by pure perceptual factors without any reference to phylogenetic development. However, the phylogenetic hypothesis remains possible since the perceptual bias acts against the phylogenetic hypothesis of better recognition of angry faces in a crowd. Nonetheless, further researchers will have to further tease apart perceptual from emotional factors in order to carefully address this question. This will in return largely improve the ecological validity of the empirical findings. We promote the use of statistical or connectionist modeling to improve the selection of faces among other stimuli, for instance by selecting angry and happy faces at equal Euclidean distances from neutral stimuli. This procedure can improve the identification of emotional versus perceptual factors in the processing of facial expression of emotion.

Acknowledgements

This work has been supported in part by the National Center for Scientific Research (CNRS UMR 6024) in addition with a grant from the French National Research Agency (ANR Grant BLAN06-2_145908, ANR Grant No. ANR-06-CORP-019 and a Clermont Université PHRC program) to Martial Mermillod and Paula M. Niedenthal. We thank Matthew N. Dailey and two anonymous reviewers for helpful and constructive comments on the manuscript.

Appendix A. First, we applied a Hann window, avoiding boundary effect in subsequent Fourier transformation. Boundary effects could result in a bias toward an over-representation of cardinal orientations, and the Hann window is a common tool to suppress this bias. The following formula describes the Hann window applied to each image

$$W(i) = 0.5 + 0.5 \times \cos\left(\frac{2\pi i}{N}\right)$$

Then, we applied Gabor receptive fields in the spectral domain by multiplying the spatial frequency information by the kernel of the Gabor function

$$G(x, y, f_c, \theta) = \frac{1}{2\pi\sigma_r\sigma_t} e^{-\frac{(x.u)^2}{2\sigma_r^2}} e^{-\frac{(x.u_\perp)^2}{2\sigma_r^2}} e^{j2\pi \underline{x} f_c}$$

With

$$\left| \begin{array}{l} \underline{x} = [x, y]^t, f_c = [f_0 \cos\theta, -f_0 \sin\theta]^t \\ u = [\cos\theta, \sin\theta]^t, u_\perp = [\sin\theta, \cos\theta]^t \end{array} \right.$$

Parameters σ_r and σ_t of the Gaussian determine the spatial extent of the filter. The vector f_c with module f_0 and direction θ describes the location of this filter in the Fourier domain.

The second component was a back-propagation neural network whose aim was to classify the output vectors provided by the Gabor filters. The connectionist network was a 3-layer back-propagation neural network. We used the standard hetero-association training algorithm.

During the feed-forward phase, activation was rescaled by means of a sigmoid transfer function:

$$f(a) = \frac{1}{1 + e^{-a}}$$

where $f(a)$ is the output activation value and a is the sum of the input activation vector multiplied by the input-to-hidden weight matrix.

The input vector activation was then propagated through the network, layer-by-layer, until it reaches the output layer. Then, the supervised learning algorithm computed the sum of squared error (SSE)

$$E = \frac{1}{2} \sum_p \sum_k (t_{pk} - O_{pk})^2$$

In this equation, p indexes the pattern in the training set, k indexes the output nodes, t_{pk} the desired output for the kth output node for the pth pattern, o_{pk} the observed output for the kth output node for the pth pattern.

Then, the error signal was computed, using the standard back-propagation algorithm (Rumelhart et al., 1986) and back-propagated through the network until the SSE function is minimized and the network "learns" the input patterns.

References

Abdi, H., Valentin, D., Edelman, B. E., & O'Toole, A. J. (1995). More about the difference between men and women: Evidence from linear neural networks and the principal component approach. *Perception, 24*, 539–562.

Dailey, M. N., & Cottrell, G. W. (1999). Organization of face and object recognition in modular neural networks. *Neural Networks, 12*(7–8), 1053–1074.

Dailey, M. N., Cottrell, G. W., Padgett, C., & Adolphs, R. (2002). EMPATH: A neural network that categorizes facial expressions. *Journal of Cognitive Neuroscience, 14*(8), 1158–1173.

De Valois, R. L., & De Valois, K. K. (1988). *Spatial vision*. New York: Oxford University Press.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3- to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General, 133*(3), 382–397.

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd – an anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917–924.

Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*, 1187–1211.

Jones, J. P., Stepnoski, A., & Palmer, L. A. (1987). The two-dimensional spectral structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology, 58*(6), 1212–1232.

Juth, P., Lundqvist, D., Karlsson, A., & Ohman, A. (2005). Looking for foes and friends: Perceptual and emotional factors when finding a face in the crowd. *Emotion, 5*(4), 379–395.

Lundqvist, D., & Litton, J. E. (1998). *The Karolinska directed faces*. Karolinska Institute.

Lyons, M. J., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with gabor wavelets. In *Proceedings of the third IEEE international conference on automatic face and gesture recognition* (pp. 200–205). Nara Japan: IEEE Computer Society.

Lyons, M. J., Budynek, J., & Akamatsu, S. (1999). Automatic classification of single facial images. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(12), 1357–1362.

Mareschal, D., French, R. M., & Quinn, P. C. (2000). A connectionist account of asymmetric category learning in early infancy. *Developmental Psychology, 36*, 635–645.

McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4*, 1–14.

Mermillod, M., Chauvin, A., & Guyader, N. (2004). Efficiency of orientation channels in the striate cortex for distributed categorization process. *Brain & Cognition, 55*(2), 352–354.

Mermillod, M., Vuilleumier, P., Peyrin, C., Alleysson, D., & Marendaz, C. (in press). The importance of low spatial frequency information for recognizing fearful facial expressions. *Connection Science*.

Mermillod, M., Guyader, N., & Chauvin, A. (2005a). Improving generalisation skills in a neural network on the basis of neurophysiological data. *Brain & Cognition, 58*(2), 246–248.

Mermillod, M., Guyader, N., & Chauvin, A. (2005b). The Coarse-to-fine Hypothesis Revisited: Evidence from neuro-computational modeling. *Brain & Cognition, 57*(2), 151–157.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality and Social Psychology, 80*(3), 381–396.
Pothos, E. M., & Close, J. (2008). One or two dimensions in spontaneous classification: A simplicity approach. *Cognition, 107*(2), 581–602.
Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception, 25*(9), 1091–1108.
Rumelhart, D. E., Hinton, G. E., & McClelland, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart, J. L. McClelland, and the PDP Research Group (Eds.), Parallel distributed processing: Explorations in the microstructure of cognition (Vol. 1). Foundations. Cambridge, MA: MIT Press.
Russell, J. A. (1994). Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin, 115*, 102–141.
Valentin, D., & Abdi, H. (1996). Can a linear autoassociator recognize faces from new orientations? *Journal of the Optical Society of America, Series A, 13*, 717–724.
Wiskott, L. (1997). Phantom faces for face analysis. *Pattern Recognition, 30*(6), 837–846.
Zhang, Z., Lyons, M., Schuster, M., & Akamatsu, S. (1998). Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perception. In *Proceedings of the third IEEE international conference on automatic face and gesture recognition* (pp. 454–459). Nara Japan: IEEE Computer Society.