Cognitive Science 33 (2009) 1413–1440

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2009.01059.x

Two Routes to Face Perception: Evidence From Psychophysics and Computational Modeling

Adrian Schwaninger,[a,b] Janek S. Lobmaier,[c] Christian Wallraven,[d] Stephan Collishaw[e]

[a]*School of Applied Psychology, University of Applied Sciences Northwestern Switzerland*
[b]*Department of Informatics, University of Zurich, Switzerland*
[c]*Department of Psychology, University of Berne, Switzerland*
[d]*Max Planck Institute for Biological Cybernetics, Tübingen, Germany*
[e]*Department of Psychological Medicine, School of Medicine, Cardiff University, UK*

Received 26 December 2006; received in revised form 6 July 2008; accepted 15 July 2008

Abstract

The aim of this study was to separately analyze the role of featural and configural face representations. Stimuli containing only featural information were created by cutting the faces into their parts and scrambling them. Stimuli only containing configural information were created by blurring the faces. Employing an old-new recognition task, the aim of Experiments 1 and 2 was to investigate whether unfamiliar faces (Exp. 1) or familiar faces (Exp. 2) can be recognized if only featural or configural information is provided. Both scrambled and blurred faces could be recognized above chance level. A further aim of Experiments 1 and 2 was to investigate whether our method of creating configural and featural stimuli is valid. Pre-activation of one form of representation did not facilitate recognition of the other, neither for unfamiliar faces (Exp. 1) nor for familiar faces (Exp. 2). This indicates a high internal validity of our method for creating configural and featural face stimuli. Experiment 3 examined whether features placed in their correct categorical relational position but with distorted metrical distances facilitated recognition of unfamiliar faces. These faces were recognized no better than the scrambled faces in Experiment 1, providing further evidence that facial features are stored independently of configural information. From these results we conclude that both featural and configural information are important to recognize a face and argue for a dual-mode hypothesis of face processing. Using the psychophysical results as motivation, we propose a computational framework that implements featural and configural processing routes using an appearance-based representation based on local features and their spatial relations. In three computational experiments (Experiments 4–6) using the same sets of stimuli, we show how this framework is able to model the psychophysical data.

Correspondence should be sent to Prof. Dr. Adrian Schwaninger, University of Applied Sciences Northwestern Switzerland, School of Applied Psychology, Institute Humans in Complex Systems, Riggenbachstrasse 16, 4600 Olten, Switzerland. E-mail: adrian.schwaninger@fhnw.ch

Keywords: Face processing; Component and configural information processing; Holistic processing; Computational modeling

1. Separate coding of featural and configural information in face perception

Faces are a complex object type, and it is surprising how well they are recognized by human beings. Even after more than 50 years a face can be recognized with 90% accuracy (Bahrick, Bahrick, & Wittlinger, 1975). Different ways have been discussed how the complex information contained in a face may be processed. Many authors have suggested that faces are processed holistically (e.g., Farah, Tanaka, & Drain, 1995; Farah, Wilson, Drain, & Tanaka, 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Various interpretations of holistic face processing have been suggested (for reviews see Maurer, Le Grand, & Mondloch, 2002; Schwaninger, Carbon, & Leder, 2003; Schwaninger, Wallraven, Cunningham, & Chiller-Glaus, 2006). The pure holistic view of face recognition claims that faces are represented as whole templates without explicitly storing the facial parts (Tanaka & Farah, 1993; see also Farah et al., 1995). Tanaka and Farah (1993) trained participants in recognizing upright faces. In the experimental phase, two faces which differed either in the shape of the eyes, nose, or mouth was simultaneously presented. In a second experimental condition the eyes, nose, or mouth was presented in isolation, that is, without the facial context. Participants had to judge which of these faces appeared in the training phase. The authors found that it was more difficult to recognize a part of a previously learned face when it was presented in isolation than when it was embedded in the facial context. This difficulty to recognize isolated parts was interpreted in favor of a holistic view of face processing, as parts do not seem to be explicitly represented.

Tanaka and Sengco (1997) hold a slightly different view of holistic face processing. They found that featural information (part-based information) and configural information are combined into holistic face representations. Whereas Tanaka and Farah (1993) and Farah et al. (1995) claimed that faces are represented as unparsed wholes without any representations of parts, the findings of Tanaka and Sengco (1997) concede that featural and configural information are first represented separately before they are integrated into a holistic representation (see also Rhodes, Brake, & Atkinson, 1993).

Maurer et al. (2002) suggest that holistic processing is one type of configural processing in which the features are ''glued together'' into a whole gestalt. According to Maurer and colleagues, configural processing refers ''to any phenomenon that involves perceiving relations among the features of a stimulus such as a face'' (Maurer et al., 2002; p. 255). This is similar to the featural-configural hypothesis postulated much earlier (e.g., Bruce, 1988; Sergent, 1984). According to Bruce (1988) configural information refers to the ''spatial interrelationship of facial features'' (p. 38), that is, the distances between features such as, for example, eyes, mouth, or nose. The spatial interrelationship of facial features was further

differentiated by Diamond and Carey (1986), who distinguished first-order and second-order relational information. First-order relational information refers to the basic arrangement of the parts (e.g., the nose lies between the eyes), whereas second-order relational information means the exact metric distances between the features. As all faces share the same first-order relational information, more importance is ascribed to second-order relational information.

In the present study we used scrambled, blurred, and intact versions of faces to investigate whether human observers only process faces holistically, or whether they encode and store the local information in facial parts (featural information) as well as their spatial relationship (configural information). These manipulations have established validity for looking at featural and configural face processing. For example, Collishaw and Hole (2000) showed that inversion had no effect on the recognition of scrambled faces, but reduced the recognition of blurred faces to chance level (see also Lobmaier & Mast, 2007). Inversion is universally accepted as predominantly affecting configural but not featural information (e.g., Bartlett & Searcy, 1993; Carey & Diamond, 1977; Diamond & Carey, 1986; Searcy & Bartlett, 1996; Sergent, 1984; for a review see Schwaninger et al., 2003; Leder & Bruce, 2000). The fact that inversion only affected blurred faces, but not scrambled faces, can be taken as evidence in favor of scrambling and blurring as manipulations to separately investigate featural and configural processing. Other authors have often separately investigated featural and configural processing by directly altering the facial features or their spatial positions. However, the effects of such manipulations are not always perfectly selective. For example, altering featural information by replacing the eyes and mouth with the ones from another face could also change their spatial relations (configural information) as mentioned by Rhodes et al. (1993). Rakover (2002) has pointed out that altering configuration by increasing the inter-eye distance could also induce a part-change, because the bridge of the nose might appear wider. Such problems were minimized in our study by using scrambling and blurring procedures that allowed investigating the role of featural and configural information separately. While scrambled faces will evidently still contain some configural information and blurring will not entirely remove featural information, these manipulations seem most appropriate for the present studies, because they reduce configural or featural information, instead of altering it.

The current study extends previous research using these manipulations (e.g., Collishaw & Hole, 2000; Davidoff & Donelly, 1990; Sergent, 1985) by ensuring that each procedure does effectively eliminate configural or featural processing. The aim of Experiments 1 and 2 was to get a clearer view of whether featural and configural representations are independent in familiar and unfamiliar face recognition. In Experiment 3 we separately scrutinized the role of first-order relational information and second-order relational information when recognizing a previously learned face. Finally, we develop a computational model that implements featural and configural processing routes using an appearance-based representation based on local features and their spatial relations. In three computational experiments (Experiments 4–6) using the same sets of stimuli as in Experiments 1–3, we show that this framework is able to provide a good model of the psychophysical data.

2. Experiment 1

Several studies have suggested that two kinds of face representations are coded in face perception: configural and featural representations (e.g., Bartlett, Searcy, & Abdi, 2003; Cabeza & Kato, 2000; Collishaw & Hole, 2000; Hayward, Rhodes, & Schwaninger, 2008; Leder & Bruce, 1998; Schwaninger, Lobmaier, & Collishaw, 2002; for an overview see Schwaninger et al., 2003, 2006). But are these representations independent of each other?

We used scrambled and blurred faces to investigate whether there is a ''transfer effect'' from featural to configural face processing, and vice versa. Testing participants in both the scrambled and blurred condition in successive blocks may reveal whether featural and configural representations are based on independent processes. If a transfer effect can be found (i.e., if the condition carried out second in Experiment 1 shows better performance), this would support the idea of interacting featural and configural representations. If, on the other hand, no effect of block order can be found, this would be consistent with two independent types of representations.

Alternatively, the performance could decrease in the condition carried out later. This would suggest that stored representations are unstable. If the shift from featural to configural processing with growing familiarity is not due to a shift in the encoding of featural and configural information, but to the growing stability of the configural representations, a decreasing performance in the blurred condition could be expected for the group tested on blurred faces after the scrambled condition.

2.1. Method

2.1.1. Participants

Twenty-four participants (12 male and 12 female) ranging in age from 20 to 46 years voluntarily took part in Experiment 1. All participants were first-year students of psychology at the University of Zurich. The participants were randomly assigned to one of two experimental groups (see below).

2.1.2. Apparatus

The experiment was run on a Windows PC using Superlab Pro 2.01. The experiment took place in a dimly lit room where participants were seated on a height-adjustable chair and responded by pressing one of five buttons on a Cedrus Response Box (RB-610). The stimuli were presented on a 17″ screen and appeared approximately 10 cm wide. A headrest ensured that the participants were at a viewing distance of 100 cm. The faces thus subtended approximately 6° horizontally.

2.1.3. Stimuli

The stimuli were created from photographs of 50 faces taken at Zurich University. Ten faces (five male, five female) were used as target faces and 40 faces (20 male, 20 female) as distractors. The faces were prepared as follows using Adobe Photoshop 6.0. The face was

extracted (i.e., without ears, neck, and hair) using the burn tool and was placed on a black background. All faces were scaled to a standard size of 300 pixels across the width of the face at pupil level. These intact faces were used in the learning phase. Fig. 1A shows an example of an intact face. The blurred stimuli were created by transforming the intact faces to black and white pictures and applying a Gaussian filter provided by Photoshop 5.5 with a radius of 8 pixels. An example stimulus is shown in Fig. 1B. Scrambled faces were cut into their parts[1] using the polygonal lasso tool with a 2-pixel feather. These parts were then scrambled in four different versions which appeared randomly. Each version was arranged so that no part was situated either in its natural position or in its natural first-order relation to its neighboring part. The parts were distributed as close to each other as possible, in order to keep the image area approximately the same size as the whole faces. An example of a scrambled stimulus can be seen in Fig. 1C.

Finally, control stimuli were created by simultaneously blurring and scrambling the parts as described above. The rationale here was that if configural and featural information is removed from a face, it will no longer be recognized above chance level. Additionally, if scrambled-blurred faces are no longer recognized above chance level this will mean that we applied sufficient blur to effectively reduce configural information. We therefore used the control stimuli to ensure that we used an appropriate blur level to reduce featural information. Fig. 1D shows an example of the control faces.

2.1.4. Task and procedure

Each participant completed four experimental conditions (blocks). Block 1 tested the recognition of intact, previously learned unfamiliar faces and was used as the baseline condition. In Block 2 recognition of blurred faces was tested, and Block 3 tested scrambled faces. In Block 4 the faces were both scrambled and blurred and served as control stimuli. Block 1 was always first, and Block 4 was always last. The order of Blocks 2 and 3 was counterbalanced across participants. Group 1 was tested with blurred faces first; Group 2 was tested with scrambled faces first.

Ten faces were chosen as target faces. None of these target faces were familiar to the participants. In each block the participants were tested on the same 10 target faces among 10 distractor faces which were different in each block. The distractor faces were counterbalanced between participants and across conditions, so that every distractor face appeared

Fig. 1. Sample stimuli. (A) Intact face, as used during the familiarizing phase and in the baseline condition of Exp. 1 and 2; (B) scrambled face; (C) blurred face; (D) scrambled-blurred face; (E) scrambled version used in Exp. 3, with the categorical relations left intact.

only once for each participant, but appeared equally often in each block over the whole experiment. The software recorded the participants' answers and reaction times.

The study phase consisted of two identical stages. The 10 target faces were successively presented for 10 s each. In the second stage the faces were presented again in the same order. After the study phase, participants were first made familiar with the test procedure. They underwent a short demonstration block, which was a shortened version of the baseline condition (three targets, three distractor faces). None of the faces used in this block were used in any of the experimental conditions. After the demonstration block, participants were asked to complete all four test conditions.

In each block participants were shown 20 faces (10 targets, 10 distractors). Each face remained visible until the participant responded. Participants were requested to respond as quickly as possible by pressing one of two keys using the left and right hand. Which hand was used for new or old faces was counterbalanced across participants. After each block, participants were able to take a short break. They could start the next block by pressing any button on the response box.

2.1.5. Analyses

A mixed-participants design was used, with condition (baseline, scrambled, blurred, control) as within-participants factor and block order as between-participants factor. Both d-prime (Green & Swets, 1966) and reaction times (RTs) were analyzed. For each condition a one-sample t test was carried out on the d-prime values in order to check the difference from chance level ($d' = 0$). A two-way analysis of variance (ANOVA) was carried out with condition (baseline, blurred, scrambled, control) as within-participants factor and block order (scrambled-blurred, blurred-scrambled) as between-participants factor.

A three-way analysis of variance (ANOVA) was run for the reaction times with condition (baseline, blurred, scrambled, control) and newness (target face, new face) as within-participants factors and block order (scrambled-blurred, blurred-scrambled) as between-participants factor.

2.2. Results

2.2.1. D-prime

The mean d-prime values were 4.0 for intact faces, 2.72 for blurred faces, 1.91 for scrambled faces, and 0.14 for scrambled-blurred faces. The one-sample t tests revealed a significant difference from 0 for intact faces, $t(23) = 25.58$, $p < .001$, blurred faces, $t(23) = 10.6$, $p < .001$, and for scrambled faces, $t(23) = 9.45$, $p < .001$ (all two-tailed). Scrambled-blurred faces were not recognized above chance, $t(23) = 0.97$, $p = .35$ (two-tailed). The ANOVA revealed a main effect of condition, $F(3, 66) = 85.69$, $MSE = 0.73$, $p < .001$. Post-hoc pairwise comparisons (Bonferoni corrected) revealed that all conditions differed significantly from each other (all $p < .001$, except for comparison blr-scr $p > .05$). The effect of block order was not significant, $F(1, 22) = 2.39$, $MSE = 1.11$, $p = .14$. The interaction between condition and block order was significant, $F(3, 66) = 3.51$, $MSE = 0.73$, $p < .05$. The results are depicted in Fig. 2.

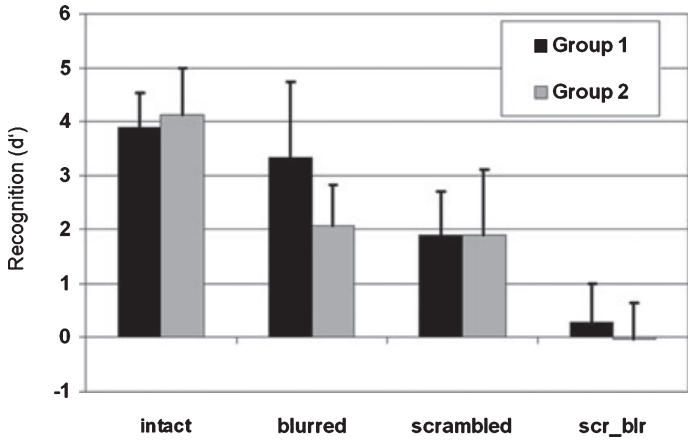

Fig. 2. Unfamiliar face recognition: D-prime values for all conditions of both groups. Group 1 was tested in the blurred condition before the scrambled condition; Group 2 carried out the scrambled condition before they were tested with blurred faces. The error bars depict standard deviations.

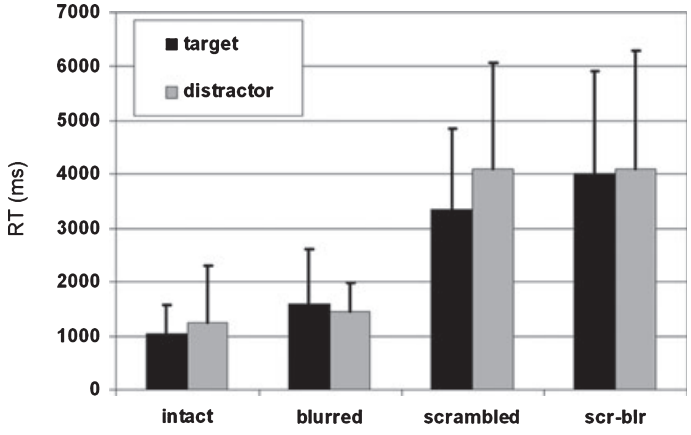

Fig. 3. Unfamiliar face recognition: reaction times for targets and distractors. The error bars depict standard deviations.

2.2.2. Reaction times

Reaction times revealed a significant effect of condition, $F(3, 60) = 37.31$, $MSE = 2{,}858{,}084.72$, $p < .001$, a significant effect of novelty, $F(1, 20) = 5.1$, $MSE = 574{,}692.19$, $p < .05$, and a significant condition × novelty interaction, $F(3, 60) = 5.65$, $MSE = 341{,}319.43$, $p < .01$. Post-hoc pair wise comparisons (Bonferoni corrected) revealed that intact faces were recognized marginally faster than blurred faces ($p = .055$), but significantly faster than scrambled faces and scrambled-blurred faces (both $p > .001$). However, scrambled faces were not recognized faster than scrambled-blurred faces. There was no effect of block order and none of the interactions with block order were significant. Therefore, RTs were pooled across groups (block order scr-blr vs. blr-scr) for calculating mean values. The mean reaction times are shown in Fig. 3.

2.3. Discussion

The first aim of the study was to assess whether participants were able to recognize unfamiliar faces on the basis of either featural or configural information. Scrambling and blurring were used to isolate each type of information. The t tests of the d-prime values revealed that faces could be reliably recognized on the basis of isolated configural and featural information, supporting a model involving two different face representations. When both types of information were eliminated (i.e., when faces were scrambled and blurred at the same time) faces were no longer recognized above chance level. The fact that scrambled-blurred faces were only recognized at chance confirms that the blurring used in this experiment effectively eliminated featural information and that scrambling eliminated configural information.

The second aim of the study was to assess whether there were any transfer effects between scrambling and blurring (i.e., facilitation for later-presented faces) and also the stability of featural and configural cues over the course of the study. Results showed no overall difference between the two groups differing in block order (scr-blr vs. blr-scr), but did reveal a significant interaction of condition × block order. This was due to the group tested on the blurred condition after the scrambled condition. This group performed less accurately in the blurred condition than the group tested in the blurred condition first. This decreasing recognition performance was not found in the scrambled condition. Why should configural memory fade while features are still remembered? A hypothesis for this effect is that featural representations are formed more easily than configural representations. In order to form reliable configural representations the faces might have to be more familiar. Diamond and Carey (1986) claim that there is a featural to configural shift in the course of expertise. Buttle and Raymond (2003) report that configural information becomes more important with growing familiarity (see also Lobmaier & Mast, 2007). Our data suggest that configural information of unfamiliar faces can only be stored for a rather short time. An alternative explanation is that processing featural information interferes with the representations of configural information: While participants were dealing with the scrambled faces the configural representations might have been weakened. However, if it is right that greater use of configural information is associated with expertise, this decrease of recognition performance should no longer be found for familiar face recognition.

Mean reaction time was shortest for the baseline condition, slightly longer for the blurred condition, longer still for the scrambled condition, and longest for the scrambled-blurred condition, explaining the main effect of condition. A target face was generally recognized faster than a distractor face was rejected, as is evident from the significant effect of newness. This could be due to identifying diagnostic characteristics in a target face. As soon as something familiar was detected, the ''target'' button might have been pressed. This was particularly the case in the scrambled condition. Participants most likely scanned every single part—to be sure that no feature was familiar—before pressing the ''new'' button. The fact that the difference of reaction times was particularly large for the scrambled condition also accounts for the significant condition × newness interaction. These results are consistent

with the assumption of a slow serial search mechanism for matching parts versus a fast parallel process for matching configural information.

3. Experiment 2

Expertise with an object class is known to enhance configural processing (e.g., Diamond & Carey, 1986; Gauthier, Skludarski, Gore, & Anderson, 2000). Also, familiarity with individual faces has been claimed to induce a shift from featural to configural processing (e.g., Buttle & Raymond, 2003; see also Lobmaier & Mast, 2007). Does configural processing gain importance in familiar faces because the configural representations are more stable? To our knowledge there is still sparse evidence on how the representation of faces changes with growing familiarity. One possibility is that a quantitative explanation accounts for face learning; that is, all aspects of the representation of the face are stored more accurately. A number of authors have now also reported that familiar faces are processed in a qualitatively different fashion than unfamiliar faces (e.g., Buttle & Raymond, 2003; Young, Hay, McWeeny, Flude, & Ellis, 1985) with evidence for the increasing importance of internal versus external facial features in familiar faces (Young et al., 1985), and an increasing sensitivity to configural changes for famous faces (Buttle & Raymond, 2003).

Our aim in Experiment 2 was to use familiar faces as target faces to further investigate the roles of featural and configural information in familiar face recognition, and to compare familiar face processing with the results of Experiment 1. Furthermore, if a lack of familiarity was the reason for the decreasing recognition performance with block order in the configural condition of Experiment 1, then this effect should not be found for familiar face recognition.

3.1. Method

3.1.1. Participants

Twenty-four participants ranging in age from 20 to 35 years took part in this experiment for course credits. All were undergraduate students of psychology at Zurich University and were familiar with the target faces. All reported normal or corrected-to-normal vision.

3.1.2. Apparatus, task, and procedure

The apparatus, task, and procedure were the same as in Experiment 1. The stimuli were also the same, but all the targets were faces of fellow students and thus familiar to the participants. The distractor faces were unfamiliar to the participants.

3.1.3. Analyses

The analyses were the same as in Experiment 1. Additionally, a two-way ANOVA was carried out on the d-prime values of Experiment 1 and 2 with familiarity (Exp 1, Exp 2) as between-participants factor and condition (intact, scrambled, blurred, scr-blr) as within-participants factor. Accordingly, a three-way ANOVA comparing the RTs of Experiment 1 and

2 was carried out with condition (intact, scrambled, blurred, scr-blr) and newness (target, new) as within-participants factors and familiarity (Exp 1, Exp 2) as between-participants factor.

3.2. Results

Mean d-prime values were 3.84 for intact faces, 3.74 for blurred faces, 2.42 for scrambled faces, and 0.43 for scrambled-blurred faces. The one-sample t test revealed a significant difference from 0 for intact faces, $t(23) = 22.91$, $p < .001$, blurred faces, $t(23) = 19.8$, $p < .001$, and for scrambled faces, $t(23) = 10.89$, $p < .001$ (all two-tailed). For scrambled-blurred faces the t test was also significant $t(23) = 2.41$, $p < .05$ (two-tailed). The ANOVA revealed a main effect of condition, $F(3, 66) = 92.9$, $MSE = 0.65$, $p < .001$. The effect of block order was not significant, $F(1, 22) = 0.001$, $MSE = 1.6$, $p = .98$. In contrast to Experiment 1, the interaction between condition and block order was not significant in Experiment 2, $F(3, 66) = 0.91$, $MSE = 0.65$, $p = .44$. The results are shown in Fig. 4.

A planned two-sample t test was carried out on the d-prime values of the baseline and the blurred condition, which revealed no significant difference between the two conditions, $t(23) = 0.38$, $p = .71$ (two-tailed).

The ANOVA comparing unfamiliar versus familiar face recognition (Experiment 1 vs. Experiment 2) revealed a significant effect of familiarity, $F(1, 46) = 6.213$, $MSE = 1.351$, $p < .05$, confirming that familiar faces were recognized more accurately than unfamiliar faces. The effect of condition remained significant, $F(3, 138) = 164.75$, $MSE = 0.732$, $p < .001$. The interaction of familiarity and condition reached statistical significance $F(3, 138) = 4.0$, $MSE = 0.732$, $p < .01$.

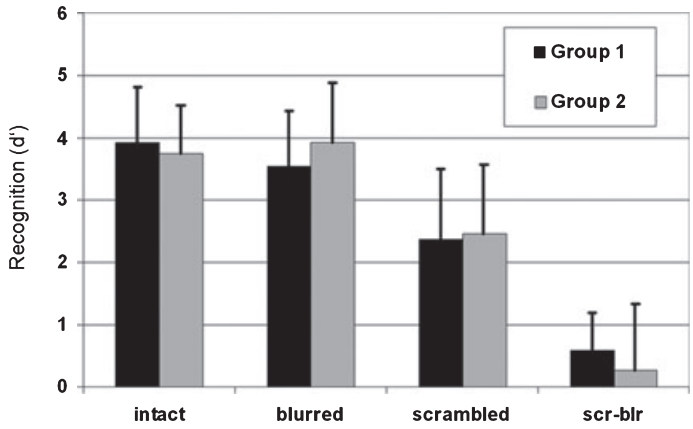

Fig. 4. Familiar face recognition: D-prime values for all conditions of both groups. Group 1 was tested in the blurred condition before the scrambled condition; Group 2 carried out the scrambled condition before they were tested with blurred faces. The error bars depict standard deviations.

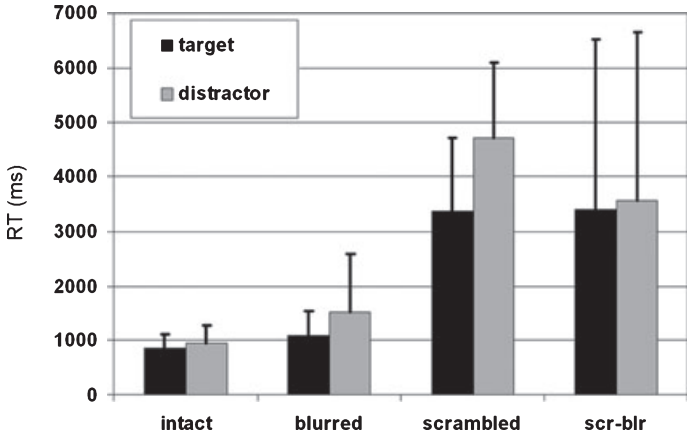

Fig. 5. Familiar face recognition: reaction times for targets and distractors. The error bars depict standard deviations.

3.2.1. Reaction times

The reaction times revealed a significant effect of condition, $F(3, 66) = 30.03$, $MSE = 3{,}859{,}114.93$, $p < .001$, a significant effect of novelty, $F(1, 22) = 20.37$, $MSE = 608{,}089.61$, $p < .001$, and a significant condition × novelty interaction, $F(3, 66) = 5.83$, $MSE = 661{,}596.84$, $p < .01$. As in Experiment 1 there was no effect of block order (blr-scr vs. scr-blr) and none of the interactions with block order were significant. Therefore, data were pooled across groups (block order scr-blr vs. blr-scr) for calculating mean RTs. The mean reaction times are shown in Fig. 5.

The ANOVA comparing the RTs of Experiment 1 and Experiment 2 revealed no significant effect of familiarity, $p = .45$, showing that unfamiliar faces were recognized just as fast as familiar faces.

3.3. Discussion

Familiar face recognition tested in Experiment 2 differs with regard to three main results from unfamiliar face recognition tested in Experiment 1. First, the overall d-prime value for blurred faces did not differ significantly from the d-prime values for the intact faces. Second, the scrambled-blurred condition was recognized slightly above chance level, and third, there was no interaction between condition and block order. More specifically, in contrast to Experiment 1, comparisons by block order in Experiment 2 showed no decrement for blurred face recognition when tested after scrambled face recognition, suggesting that configural representations are more stable and robust for familiar faces.

The high-recognition performance of the blurred faces supports the idea that configural processing becomes more accurate and stable when faces are familiar. An overall main effect showed that familiar face recognition was more accurate in terms of encoding and storing featural and configural information (quantitative difference). In

addition, there is also a qualitative difference when the stability of the encoding is considered. While configural representations appeared to fade over time for unfamiliar faces (Experiment 1), they were more stable when faces were familiar. The fact that there was no decrease of d-prime value in the blurred condition of Experiment 2 as opposed to Experiment 1 is consistent with the hypothesis that with growing familiarity configural information is remembered better. If faces are unfamiliar it is much more difficult to remember the configuration of a face. Familiar faces, on the other hand, have been encountered much more often and therefore there is no decrease in recognition performance for configural information.

In the control condition the scrambled-blurred faces were recognized slightly above chance level; post-hoc analyses revealed that this was only due to one participant group (block order blr-scr). Moreover, it is important to note that performance in this group was only slightly above chance and considerably worse than when faces were only scrambled or only blurred. In the other group (scr-blr) performance was at chance in the scrambled-blurred condition. Taken together, these results suggest that configural and featural processing was substantially impaired by scrambling and blurring. The fact that there was no condition $\times$ block order interaction supports the view that the two processes work independently; no transfer effect was found from either condition to the other.

Regarding reaction times, the significant effect of condition once again reflects the difficulty of the task. The baseline and the blurred condition both reveal very short reaction times, whereas the reaction times of the scrambled and scrambled-blurred conditions were rather long. This difference may reflect the cognitive processes underlying face recognition. Both the intact faces and the blurred faces could be processed holistically, whereas for the scrambled condition RT data seem to be more consistent with a slower serial search mechanism in which parts are processed separately in order to match them to memory representations. The shorter RTs for target faces further accounts for this claim.

In summary, the data of Experiment 2 confirmed and extended findings of Experiment 1. As in the previous experiment both scrambled and blurred faces were recognized at above chance levels, indicating that both featural and configural representations can be used independently of one another to recognize faces. Scrambled-blurred faces were processed at chance (group blr-scr) or just above chance level (group blr-scr), indicating that together the two manipulations eliminated most or all of the featural and configural cues in the stimuli. Familiar faces were on the whole recognized more accurately than unfamiliar faces, reflecting a quantitative advantage with growing familiarity. In addition, there was a significant interaction between familiarity and condition indicating a shift towards configural processing for familiar faces. In fact, blurred familiar faces were recognized as accurately as intact faces, even when they followed a block of intervening scrambled faces, suggesting that configural representations are more stable and robust for familiar faces. In line with a recent study by Lobmaier and Mast (2007), the present data suggest a difference in processing, namely that configural face representations of familiar faces are processed more accurately than those of unfamiliar faces.

4. Experiment 3

Diamond and Carey (1986) distinguished first- and second-order relational information. Second-order relational information is defined as the exact distances between the features, while first-order relational information describes the relative position of a feature in the face. The terms ''metric spatial relations'' and ''categorical spatial relations'' (Kosslyn, 1994) make a similar distinction. The scrambling used in Experiments 1 and 2 destroyed both first- and second-order relational information. Nothing can be said about the spatial dependence of featural representations. The aim of Experiment 3 was to scrutinize whether featural representations are indeed independent of both first- and second-order relational information. In the scrambled faces of Experiment 3 we left the categorical relations intact but changed the metrical distances between the parts. If categorical spatial relations are explicitly represented, a face would be better recognized when the features are left in their categorical spatial relations. On the other hand, featural representations may be relatively independent of their spatial relationship both in terms of first- and second-order relational information. In this case we would expect no increase of sensitivity when the parts are left in their categorically correct location.

4.1. Method

4.1.1. Participants

Twelve undergraduate students of Zurich University ranging in age between 20 and 35 years voluntarily took part in Experiment 3. All participants were naïve to the aim of the study and did not take part in any other experiments reported here. All reported normal or corrected-to-normal vision and were unfamiliar with all the test faces.

4.1.2. Apparatus

The apparatus was the same as in the previous experiments.

4.1.3. Stimuli

The same intact and blurred stimuli were used as in Experiments 1 and 2. New scrambled stimuli were created by placing the parts in their categorically correct position, but destroying the precise metric spatial relations (categorical scr). The same parts were used as in Experiments 1 and 2. An example stimulus is shown in Fig. 1E.

4.1.4. Task and procedure

The task and procedure were comparable to that used for Group 2 in Experiment 1. Participants were tested in the baseline condition with intact faces, then with the new scrambled faces, and finally with blurred faces. Scrambled-blurred faces were not tested in this experiment. The results of the categorical scrambled faces could then be directly compared with the results in the scrambling condition of Group 2 in Experiment 1.

4.1.5. Analyses

As in the previous experiments the d-prime value was calculated for each participant. A one sample t test was carried out on the d-prime values of each condition in order to check the difference from chance level. Then a two-sample t test was carried out comparing the results of the scrambled condition of Group 2 in Experiment 1 and the scrambled condition in this experiment. A two-way analysis of variance (ANOVA) was additionally carried out with condition (base, scr, blr) as within-participants factor and group (metric scr, categorical scr) as between-participants factor.

4.2. Results

The mean d-prime values were 4.06 for intact faces, 1.73 for scrambled faces, and 2.16 for blurred faces. The one-sample t test revealed a significant difference from 0 for intact faces, $t(11) = 21.9$, $p < .001$, scrambled faces, $t(11) = 6.05$, $p < .001$ and blurred faces, $t(11) = 9.84$, $p < .001$ (all two-tailed). The two-sample t test revealed no significant difference between the two scrambling conditions of Experiment 3 and Group 1 in Experiment 1, $t(22) = 0.4$, $p = .70$ (two-tailed). The two-way ANOVA revealed only a significant effect of condition, $F(2, 44) = 60.49$, $MSE = 0.61$, $p < .001$. There was no effect of group (metric scr, categorical scr) and no two-way interaction condition × group, $F(2, 44) = 0.09$, $MSE = 0.51$, $p = .76$. The results are shown in Fig. 6.

4.3. Discussion

Theories of face perception highlight distinctions between different types of configural processing. One important distinction is between first-order categorical relationships specifying that a stimulus is a face and second-order relational cues that vary between faces

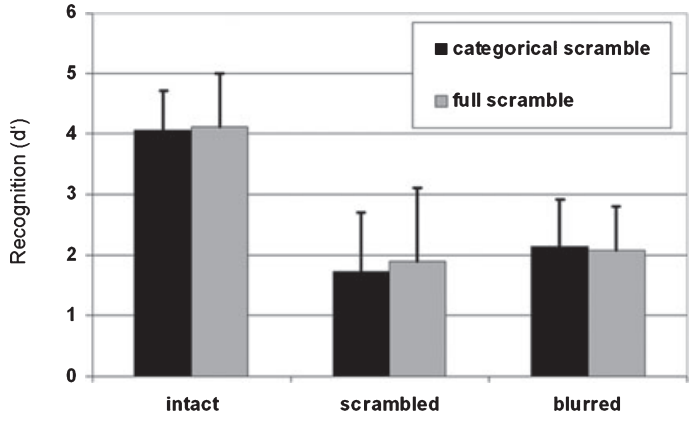

Fig. 6. D-prime values of scrambled and blurred condition. Dark bars depict the group tested in categorical scrambling condition (i.e., where categorical spatial relations are left intact, Categorical Scr), and gray bars depict values of Group 2 in Exp. 1 (Full Scramble). The error bars depict standard deviations.

(Diamond & Carey, 1986; Maurer et al., 2002). According to these theories, second-order relational information should be of greater importance in recognizing individual faces, but there has been little or no research that has tested whether first-order categorical information makes some additional contribution to face recognition, either alone, or in interaction with featural or configural cues. In Experiment 3 faces were cut into parts, which were placed in their categorically correct place. These types of scrambled faces were not recognized more accurately than faces whose parts have been scrambled and placed in their categorically incorrect place (scramble condition in Experiment 1). These findings support the view that the explicit representation of categorical relations is of no use for identifying faces. Featural representations seem to be independent of the spatial arrangements of the facial parts. More specifically, facial features are processed regardless of their spatial location; the use of featural information in face recognition seems to be independent of its location.

5. Experiments 4–6

Experiments 1–3 showed that human beings can independently process featural and configural face information and that these two information types constitute two separate routes to face processing. Experiment 3 further showed that featural information is independent of its spatial location in the face. In the following we aim to design and test a computational implementation of the two-route processing model for face recognition (Experiments 4–6). Research on face recognition in the context of computer vision can be roughly divided into three areas:

1. Feature-based approaches process an image of a face to extract features—these can range from simple, high-contrast features to high-level, semantic facial features.
2. Holistic approaches use the full image pixel information of the face image.
3. Hybrid systems combining these two approaches.

The earliest work in face recognition focused almost exclusively on high-level, feature-based approaches. Starting in the 1970s, several systems were proposed that relied on extracting facial features (eyes, mouth, and nose) and in a second step calculating two-dimensional geometric properties of these features (Kanade, 1973). Although it was shown that recognition using only geometric information (such as distances between the eyes, the mouth, etc.) was computationally effective and efficient, the robust, automatic extraction of such high-level facial features has proven to be very difficult under general viewing conditions (Brunelli & Poggio, 1993). One of the most successful face recognition systems based on local image information therefore used much simpler features based on Gabor-filter responses, which are collected over various scales and rotations and then processed using a complex, graph-based matching algorithm (Wiskott, Fellous, Krüger, & v. d. Malsburg, 1997). The advantage of such low-level features lies in their conceptual simplicity and compactness.

In the early 1990s, Turk and Pentland (1991) developed a holistic recognition system called ''Eigenfaces,'' which used the full pixel information to construct an appearance-based, low-dimensional representation of faces—a face space. This general idea of a face space is shared by other algorithms such as Linear Discriminant Analysis (LDA; Belhumeur, Hespanha, & Kriegman, 1997), Independent Component Analysis (ICA; Bartlett, Movellan, & Sejnowski, 2002), Non-negative Matrix Factorization (NMF; Lee & Seung, 1999), or Support Vector Machines (SVMs; Phillips, 1999). The main difference between these algorithms lies in the statistical description of the data as well as in the metrics used to compare different elements of the face space. The advantage of PCA (and other holistic approaches) in particular is that they develop a generative model of facial appearance that enables them, for example, to reconstruct the appearance of a noisy or occluded input face. An extreme example of this is the morphable model by Blanz and Vetter (for recognition applications, see Blanz & Vetter, 2003 and Weyrauch, Heisele, Huang, & Blanz, 2004), which does not work on image pixels but on three-dimensional data of laser scans of faces. Because of their holistic nature, however, all of these approaches require specially prepared training and testing data with very carefully aligned faces in order to work optimally.

Given the distinction between local and holistic approaches, it seems natural to combine the two into hybrid recognition architectures. Eigenfaces can of course be extended to ''Eigenfeatures'' by training facial features instead of whole images. Indeed, such systems have been shown to work much better under severe changes of the appearance of the face such as due to occlusion by other objects or make-up (see Swets & Weng, 1996). Another system uses local information extracted from the face to fit a holistic shape model to the face. For recognition, not only holistic information is used but also local information from the contour of the face (Cootes, Edwards, & Taylor, 2001). Finally, in a system proposed by Heisele, Ho, Wu, and Poggio (2003), several SVMs are trained to recognize facial features in an image, which are then combined into a configuration of features by a higher-level classification scheme. Again, such a scheme has been shown to outperform other, purely holistic, approaches.

Recently, there has been growing interest in testing the biological and behavioral plausibility of some of these approaches (e.g., Furl, O'Toole, & Phillips, 2002; Riesenhuber, Jarudi, Gilad, & Sinha, 2004; Schwaninger, Wallraven, & Bülthoff, 2004; Wallraven, Schwaninger, & Bülthoff, 2004, 2005). In Schwaninger et al. (2004) and Wallraven et al. (2004, 2005) we proposed a simple, computational implementation of the two-route processing described in this paper and showed that it could capture the psychophysical data on face recognition obtained by Schwaninger et al. (2002). The computational model was designed using a low-level, feature-based face representation consisting of salient image features that were extracted at a detailed and a coarse image scale. The detailed image features were used for the component route, whereas the configural route was modeled using the coarse image features *and* their spatial layout.

The aims of Experiments 4–6 are to extend our previous results (Schwaninger et al., 2004; Wallraven et al., 2004, 2005) by modeling the psychophysical data on configural and component processing obtained in Experiments 1–3, respectively.

In addition, we will compare and discuss the proposed computational model in the context of other feature-based and holistic models.

5.1. Computational implementation of component and configural processing

In the following we describe the computational implementation, which is largely based on Wallraven et al. (2004, 2005). The core question that this implementation tries to address is how to formulate configural and component information algorithmically so that they become amenable to computational modeling. For this, we use two basic ingredients: the data representation, which in our case amounts to specifying how to extract appearance and spatial features from visual input, and the data processing, which in our case consists of the algorithms which manipulate the representations in order to match, for example, a new face to an old face.

The implementation uses an appearance-based representation based on local features that are extracted at two image scales. In this context, ''appearance-based'' means that the representation is directly derived from the visual input. The representation is based on the concept of ''local features,'' which can be defined as robustly localizable subparts of an image—examples for these kinds of local features range from low-level features such as regions of high changes in image intensity to higher-level features such as eyes, mouth, and nose in the case of faces. The reason for choosing local features rather than global, holistic ones (see also discussion above) lies in their increased robustness to changes in viewing conditions such as occlusion, lighting, etc. Finally, our implementation uses a multi-scale approach by analyzing image content at multiple spatial frequencies. The main reason for this is that as was shown in earlier studies, configural and component information seem to be extracted and processed at different spatial frequency scales (Goffaux, Hault, Michel, Vuong, & Rossion, 2005). The frequency ranges of the two scales in our implementation therefore correspond closely to the ones that were found to be important for the processing of component (>32 cycles per face width) and configural (<8 cycles per face width) information in their study.

More specifically, given an image of a face, it is first low-pass filtered to obtain the two image scales. On each scale, the image is processed by a Harris corner detector (Harris & Stephens, 1988), which extracts salient image locations in the image based on the strength of local image intensity gradient. The appearance-based feature information then simply consists of a small image patch (5×5 pixels) that is extracted around each feature location in the image. The spatial feature information is determined by its *embedding*, which consists of a vector containing two-dimensional pixel distances to a number of neighboring features. The number of features to which the distance is evaluated varies for the component or the configural properties of each feature: For component information, a local, small neighborhood is used, whereas for configural information, a global, large neighborhood is specified. The extent of the neighborhood for each of the two scales constitutes a free parameter of the system—in this study, however, the two parameters are fixed.

Fig. 7 shows a reconstruction of a face from such a feature representation. Note how despite the fact that the feature extraction algorithm is not designed for detecting facial features, the features tend to cluster around semantically important facial features, such as the

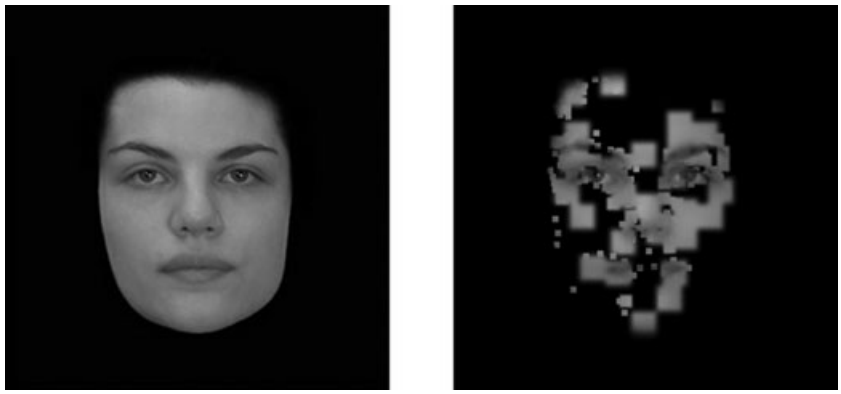

Fig. 7. Original face (left) and reconstruction from its feature representation (right). Blurred features originate from the coarse scale, whereas detailed features originate from the fine scale. Note how features tend to cluster around facial landmarks (eyes, mouth).

eyes, nose, corners of the mouth, etc. The total number of features that are extracted at each scale is an additional input parameter to the implementation—the reconstruction shown in Fig. 7 uses 25 features at each scale.

The second major component of the computational modeling effort consists of specifying a matching algorithm that can recognize faces, that is, to determine whether an image of a face has previously been seen. As each image consists of a set of local features (consisting of appearance-based image patches and spatial embeddings), recognition amounts to finding the *best matching feature set* between a test image and all learned images. The two routes for face processing in this case are implemented with two different matching algorithms based on configural and component information. Each feature is matched to all other features in an image using two terms: The first term specifies the *appearance* similarity of the two image patches, which is done by calculating a normalized cross-correlation between the image intensity values in the two patches. The second term determines the *geometric* similarity between the embeddings of the features. This is done by evaluating the Euclidean distance between the two embedding vectors. To reiterate, *component* matching is done on the higher-frequency scale using a *local-neighborhood* analysis, whereas *configural* matching is done on the lower-frequency scale using the *global* neighborhood relations between features. In a final step, we then determine a one-to-one mapping between all features of the source image to the target image. The percentage of matches for the component route *and* the configural route between two images then constitutes two matching scores, which averaged together yield the final matching score.

5.2. Experiment 4—Recognition of scrambled and blurred faces

5.2.1. Stimuli

In order to compare the computational results with the psychophysical data, the computational experiments used the same sets of stimuli as the studies conducted in Experiment 1–3.

5.2.2. Task and procedure

Each of the 10 target images was first encoded yielding the local feature representation and its configural information. In a simulated old-new experiment, the 10 target images as well as 10 distractors were then presented to the system in the blurred, scrambled, and scrambled-blurred conditions. Each image was encoded and, using the feature matching algorithm, matched against the previously learned images, which resulted in 10 matching scores.

In a next step, the scores were converted into a performance measure that can be directly compared with the psychophysical data. For this, they were converted into an ROC-curve by thresholding the matching scores for the target faces (resulting in hit-rates as a function of the threshold) as well as the matching scores for the distractor faces (resulting in false-alarm-rates as a function of the threshold). Finally, the area under the ROC-curve was measured (this measure is abbreviated as AUC in the following) yielding a nonparametric measure of recognition performance ($0.5 \leq AUC \leq 1.0$). This procedure was repeated 10 times with different subsets of target and distractor faces in order to be able to statistically analyze *variations* in the computational recognition results. Similarly, the human d′-scores were converted to AUC scores (see Green & Swets, 1966).

Furthermore, we ran the same experiments with three additional computational algorithms. The first algorithm used the same matching strategy, albeit without the geometric term, which allowed us to assess the advantage gained by adding spatial layout information for feature matching in the configural route. The second algorithm is a state-of-the-art local feature framework based on scale-invariant features (SIFT, Lowe, 2004) that was shown to provide excellent performance in a number of object recognition tasks. Local features in this framework consist of scale-invariant, high-dimensional (each feature vector has 128 dimensions) histograms of image gradients at local intensity maxima. The SIFT algorithm is available for download at http://cs.spider.uk.ca/~lowe/ and was used without modification in the following experiment. Finally, we wanted to compare modeling performance to a simple holistic matching algorithm. For this third algorithm, the image representations simply consisted of all image pixels of the face images. Matching was done by considering the image pixels as a vector and then simply evaluating the Euclidean distance between two pixel vectors.

5.2.3. Results and discussion

Fig. 8 compares AUC-values for human data with AUC-values for the computational implementation. In addition, the computational data are separated to show the contributions of the configural route and the component route in the different conditions. First, it can be seen that the computational performance is slightly lower than the human performance. This can be attributed to the simple visual features that were used in our implementation. More importantly, however, the relative contribution of the two processing routes follows exactly the expected pattern with the configural route being active in the blurred condition and the component route being active in the scrambled condition. In addition, the configural route does not contribute to recognition in the scrambled condition; similarly, the performance of the component route in the blurred condition is negligible. Performance of both routes

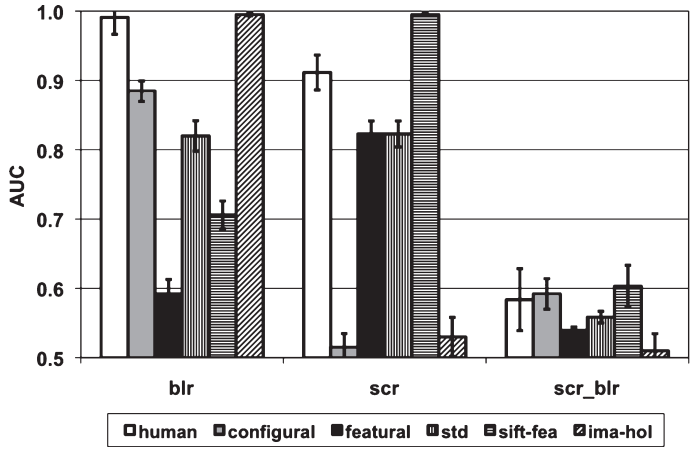

Fig. 8. Unfamiliar face recognition: AUC values for all conditions of the human data as well as the computational modeling data. The computational data is split into contributions of the component and configural processing route, as well as standard local feature matching (std), matching with SIFT features (sift-fea), and with holistic image representations (ima-hol). Performance for intact faces is at AUC = 1.0 and is not shown here. In addition, performance for the combined model does not differ from the single routes for each condition. All error bars depict SEM.

reaches chance level in the scrambled and blurred condition. In addition, the relative contributions of each route closely follow the human data.

Taken together, this pattern of results models the psychophysical experiments on a qualitative level and thus provides initial evidence for the perceptual plausibility of our implementation of the two routes of visual processing. In Fig. 9, an example of feature matching in each of the three conditions is given—corresponding features are indicated as white dots. In this example, the component route is active for the scrambled condition, the configural route for the blurred condition, whereas only one match could be found in the scrambled and blurred condition. The full experimental results in Fig. 8 confirm that both routes process the information independently as AUC-values are negligibly small for the conditions in which only one type of information should be present. In addition to the quantitative results and the relative activation of the two routes in the different condition, this provides further evidence for the plausibility of the implementation.

Fig. 8 also shows the results of standard local feature matching *without* the geometric constraint on the stimuli. Whereas there is no difference for the scrambled stimuli (which is not surprising, given that both algorithms are virtually identical), recognition performance in the blurred condition drops to the level of performance in the scrambled condition. This result demonstrates that the additional geometric constraint not only helps to increase recognition performance but that this local feature-matching framework seems necessary to capture the performance pattern observed in the human data.

Performance for the feature-based SIFT approach (sift-fea) is rather poor in the blurred condition, whereas the scrambled condition yields almost perfect recognition rates and the scrambled-blurred condition drops to chance levels. The inferior performance in the blurred

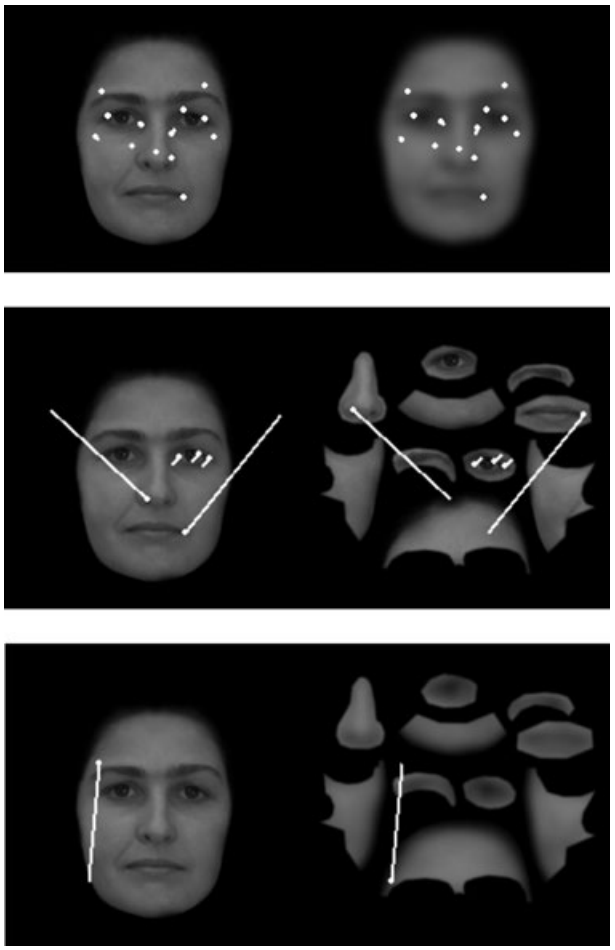

Fig. 9. Corresponding features for the three test conditions: The left face in all three rows shows a learned training face, whereas the right face is from one of the three test conditions (upper row: blurred face; middle row: scrambled face; lower row: scrambled and blurred face). The lines on both the training as well as the testing faces connect the corresponding features in the image plane, respectively. In the blurred condition, the only matches stem from the configural route; in the scrambled condition, only the featural route is active. The one false match shown here in the scrambled and blurred condition is due to a featural match.

condition is due to the fact that only a few SIFT features are extracted in the blurred images, which severely limits the discriminatory power of the feature matching. In the scrambled condition, however, the many detailed, high-dimensional features that are extracted for both scrambled and intact images guarantee a high degree of recognition performance.

Finally, as can be seen in Fig. 8, the holistic approach (ima-hol) shows the exact opposite pattern to the feature-based approach: almost perfect recognition performance in the blurred condition with chance performance in both the scrambled and scrambled-blurred condition. This pattern is not surprising given that the coarse outline of the pixel information in the

intact images is preserved in the blurred condition. It is perhaps interesting to note that such a pattern of performance would also be predicted for the Elastic Bunch Graph Matching method proposed by Wiskott et al. (1997). Even though this algorithm has been shown to be able to tolerate substantial changes in viewing conditions (such as changes in lighting, image plane rotation, as well as some invariance to rotation in depth and moderate degrees of occlusion), a fully scrambled face as used here and in the perceptual experiments would not be recognizable anymore in this approach.

In summary, none of the other computational approaches is capable of modeling the relative contribution of the component and configural route on its own—both the failure of purely feature-based and purely holistic approaches speak strongly in favor of a hybrid approach integrating appearance-based and configural information.

5.3. Experiment 5—Effect of familiarity

In a second step it was tested how well the computational implementation would be able to capture the effects of familiarity observed in the psychophysical experiments. One of the most obvious parameters that might be responsible for the difference between familiar and unfamiliar face recognition might be the richness or complexity of the extracted representation. If humans are repeatedly exposed to the same face, this experience could simply result in a more detailed representation of its visual appearance. The computational counterpart to this in our computational implementation would consist of the number of local features that constitute the representation of a face image. The following computational experiment explicitly tested this hypothesis with the stimulus set of the previous experiment by systematically increasing the number of features in each processing route.

5.3.1. Results and discussion

Fig. 10 shows AUC-values for the human data from Experiment 2 compared with AUC-values for the computational implementation. The computational data are shown for three different sizes of the visual representation: original (same as in the previous experiment), the number of local features increased by 50%, and the number of features increased by 10%. As hypothesized, the performance of the computational data increases with increasing visual complexity in both routes. In contrast, the results for the configural route in the scrambled condition and for the component route in the blurred condition show no systematic increase with increasing visual complexity. Most importantly, the relative contribution of each route does not change in the three conditions. In addition, the performance of the most complex visual representation approaches human performance—a further increase in number of features, however, does not provide better recognition performance, indicating that the discriminatory power of the simple visual features used in this study has reached its limits. The experimental results presented here suggest that a surprisingly simple parameter such as the complexity of the visual representation might be sufficient to explain the increase in performance observed in the psychophysical experiments.

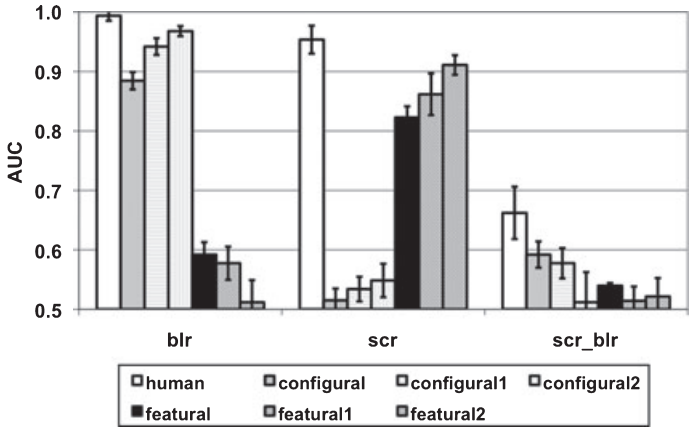

Fig. 10. Familiar face recognition: AUC values for all conditions of the human data as well as the computational data split into contributions by the component (featural) and configural processing route. Computational data are based on three visual representations with increasing visual complexity. The error bars depict *SEM*.

5.4. Experiment 6—Effect of scrambling type

Whereas in the previous two computational experiments we were interested in modeling unfamiliar and familiar face recognition, in this experiment we wanted to reproduce the independence of scrambling type found in Experiment 3 in the original psychophysical study. The computational experiment was therefore repeated with the same set of categorically scrambled stimuli and compared with the results from the non–categorically scrambled face images used before.

5.4.1. Results and discussion

The results of this computational experiment are shown in Fig. 11 for the two types of scrambling (Cat and Tot). Similarly to the human data, the computational performance remains unaffected by type of scrambling used, thus providing further support for the plausibility of our implementation. This is confirmed by a two-sample t test (two-tailed), which yields no significant difference between the two conditions for the component processing route, $M = 0.82$, $t(11) = 1.46$, $p = .16$.

6. General discussion

In this study we investigated the role of featural and configural representations in familiar and unfamiliar face recognition. In three psychophysical experiments, featural and configural information was presented in isolation, testing whether faces could still be recognized on the basis of only one kind of information. All three experiments support a face processing model that includes separate configural and featural representations. Many authors have argued that upright faces are processed holistically (Biederman & Kalocsai, 1997; Farah

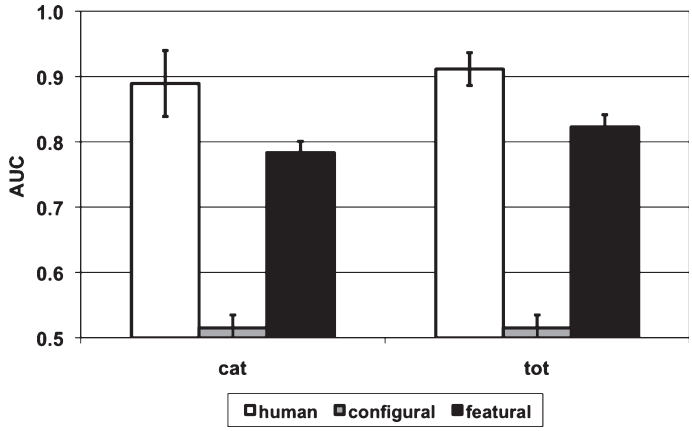

Fig. 11. AUC values of scrambling condition for human and computational data where categorical spatial relations are left intact (Cat) versus where they are totally scrambled (Tot). The error bars depict *SEM*.

et al., 1995; Tanaka & Farah, 1993). There are different ways of defining holistic processing (for reviews see Maurer et al., 2002; Schwaninger et al., 2003, 2006). A purely holistic view of face processing in which featural information is not explicitly represented is inconsistent with our results. In all experiments faces were reliably recognized when only featural information was provided (scrambled condition). This is consistent with the model proposed by Schwaninger et al. (2002, 2003), according to which faces are first represented in the primary visual areas as pictorial metric input representations. From these input representations specific information is extracted in order to form featural and configural representations. The output of these representations then converges to the same face identification units, which integrate featural and configural information to ''holistic'' representations. Note that this understanding of holistic differs from the original concept formulated by Tanaka and Farah (1993) and Farah et al. (1995) who claim that parts (featural information) are not explicitly represented. Our data clearly suggest a dual-code view where featural and configural information is represented separately before it is combined into a holistic face representation. This is very much in accordance with findings of several other studies (e.g., Bartlett et al., 2003; Cabeza & Kato, 2000; Collishaw & Hole, 2000; Rhodes et al., 1993; Tanaka & Sengco, 1997). These assumptions are based on behavioral data and it is certainly interesting to compare our data with results from cognitive neuroscience. In their review, Rossion and Gauthier (2002) remark that no current fMRI or anatomical data give evidence that facial features are extracted before they are combined to a holistic representation. Yet, Haxby, Hoffman, and Gobbini (2000) suggest that a region of the inferior occipital gyrus may be involved in the perception of facial parts. It will have to be the aim of future work to repeat these experiments with methods of cognitive neuroscience, in order to find out whether featural processing can be anatomically dissociated from configural processing.

Our study showed that there is no transfer effect in terms of a performance increase from blurred to scrambled recognition and vice versa, which is consistent with the assumption of separate representations for featural and configural information. Moreover, the results of

Experiments 1 and 2 suggest that both featural and configural representations are used for familiar and unfamiliar face recognition. Familiar faces were recognized more accurately than unfamiliar faces and the comparisons by block order in Experiment 1 and Experiment 2 indicated that configural representations are more stable and robust for familiar faces.

Experiment 3 revealed that featural representations seem to be independent of the spatial arrangement of the facial parts. The fact that first- order relational information did not increase recognition accuracy is consistent with the assumption that featural representations are independent of first- and second-order relational information. This leads to the conclusion that categorical relational information is not crucial for recognizing individual faces. Note, however, that representations of categorical relations may be important for recognizing that a stimulus is a face as suggested by Maurer et al. (2002) (see also Diamond & Carey, 1986).

As a second focus of our work, we have presented a computational framework based on local features and their spatial relations that was motivated by these two hypothesized routes of facial processing. In summary, our results show that our implementation of the two-route architecture is able to capture the range of human performance observed in the psychophysical experiments. In addition, changes in the internal parameters of the architecture—we have so far investigated visual complexity and discriminability—result in plausible changes in observed performance while retaining the overall qualitative similarity to the human data in terms of the observed weighting of the two routes.

In the following, we discuss certain aspects of the proposed architecture in more detail. As we have seen, by using more discriminative features it becomes possible to achieve very good recognition performance for scrambled images, whereas by using holistic approaches, performance is very good for blurred images. Whereas one might be able to model human performance by a suitable combination of those two approaches, our implementation offers an integrated, more parsimonious framework for recognition rather than postulating two very different face representations. From a cognitive perspective, in addition, there is evidence that humans do not seem to pay attention to images at the level of single pixel information, as would be the case for the holistic computer vision techniques. Humans rather seem to rely on a more abstract representation of visual data maybe even including a semantic representation such as ‘‘full mouth,’’ ‘‘curved eyebrows,’’ which is based on a higher-level interpretation of the visual information. Although our proposed computational model is not semantically grounded, it is extendable to a semantic and thus class-specific representation (see, e.g., Ullman, Vidal-Naquet, & Sali, 2002 for an approach in this direction).

Alternatively, one might also base the need for a more abstract representation simply on memory or storage constraints: The amount of visual memory necessary to save holistic, detailed pixel information is simply not available for this task. The proposed implementation of the two processing routes can be seen as an embodiment of such a memory constraint: The huge number of possible visual features and their image relations is reduced to a few of the most salient ones taking into account their local neighborhood for a larger number of detailed features and their global neighborhood for a smaller number of coarse features. Whereas from a computer vision perspective the task itself could be solved with almost perfect recognition performance—even though at a significantly higher memory

load—the extraction of visual features enables a much sparser and more abstract representation. In addition, their inherent robustness allows for extraction of further abstract information—such as analysis of visual features across all learned faces to extract parts and common feature relations, etc. Apart from providing one layer of data abstraction, our implementation of the visual features underlying the two processing routes thus seems to be able to fit well into models of human visual memory.

In this context, it is important to stress that our focus in the implementation of the computational model has not been on developing efficient, low-level features for face recognition. Indeed, it is easily possible to integrate state-of-the-art features such as SIFT (Lowe, 2004) or Gabor Jets (Wiskott et al., 1997) into the configural and featural processing pipeline. Nevertheless, in order to claim more generality and applicability in the domain of face recognition, the model would of course need to be tested with other recognition tasks, such as generalization across view and illumination changes, sensitivity against occlusions, as well as dealing with facial expressions. Providing these tests alongside with further improvements in the algorithm is our current topic of research.

In summary, Experiments 1–3 have provided converging evidence for the view that component and configural information are processed separately, encoded explicitly, and used automatically in familiar and unfamiliar face recognition. A computational model that specifies the processes and representations has been developed. The computational Experiments 4–6 have shown that this model is psychophysically very plausible since very similar results were obtained as in the psychophysical Experiments 1–3.

Note

1. In order to identify the parts of a face, a free-listing experiment was run with 41 students. The most frequently named parts (named by more than 80% of the participants) were as follows: nose, eyes, cheeks, forehead, eyebrows, chin, ears, and mouth (listed by frequency). The ears were excluded for technical reasons, leaving a total of 10 parts to be scrambled.

References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General*, *104*, 54–75.

Bartlett, J. C., & Searcy, J. (1993). Inversion and configuration of faces. *Cognitive Psychology*, *25*(3), 281–316.

Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, *13*(6), 1450–1464.

Bartlett, J. C., Searcy, J. H., & Abdi, H. (2003). What are the routes to face recognition? In M. A. Peterson & G. Rhodes (Eds.), *Perception of faces, objects and scenes: Analytic and holistic processes* (pp. 21–52). New York: Oxford University Press.

Belhumeur, P., Hespanha, J., & Kriegman, D. (1997). Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(7), 711–720.

Biederman, I., & Kalocsai, P. (1997). Neurocomputational bases of object and face recognition. *Philosophical Transactions of the Royal Society of London, B*, *352*, 1203–1219.

Blanz, V., & Vetter, T. (2003). Face recognition based on fitting a 3D morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *25*, 1063–1074.

Bruce, V. (1988). *Recognising faces*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Brunelli, R., & Poggio, T. (1993). Face recognition: Features versus templates. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, *15*(10), 1042–1062.

Buttle, H., & Raymond, J. E. (2003). High familiarity enhances visual change detection for face stimuli. *Perception & Psychophysics*, *65*(8), 1296–1306.

Cabeza, R., & Kato, T. (2000). Features are also important: Contributions of feutural and configural processing to face recognition. *Psychological Science*, *11*(5), 429–433.

Carey, S., & Diamond, R. (1977). From piecemal to configurational representation of faces. *Science*, *195*, 312–314.

Collishaw, S. M., & Hole, G. J. (2000). Featural and configurational processes in the recognition of faces of different familiarity. *Perception*, *29*, 893–910.

Cootes, T., Edwards, G., & Taylor, C. (2001). Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *23*, 681–685.

Davidoff, J., & Donelly, N. (1990). Object superiority: A comparison of complete and part probes. *Acta Psychologica*, *73*, 225–243.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General*, *115*, 107–117.

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What causes the face inversion effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 628–634.

Farah, M. J., Wilson, K. D., Drain, M., & Tanaka, J. N. (1998). What is ''special'' about face perception? *Psychological Review*, *105*(3), 482–498.

Furl, N., O'Toole, A. J., & Phillips, P. J. (2002). Face recognition algorithms as models of the other race effect. *Cognitive Science*, *96*, 1–19.

Gauthier, I., Skludarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197.

Goffaux, V., Hault, B., Michel, C., Vuong, Q. C., & Rossion, B. (2005). The respective role of low and high spatial frequencies in supporting configural and featural processing of faces. *Perception*, *34*, 77–86.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Harris, C., & Stephens, M. J. (1988). A combined corner and edge detector. *Proceedings of the Alvey Vision Conference*, 147–152.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, *4*, 223–233.

Hayward, W. G., Rhodes, G., & Schwaninger, A. (2008). An own-race advantage for components as well as configurations in face recognition. *Cognition*, *106*, 1017–1027.

Heisele, B., Ho, P., Wu, J., & Poggio, T. (2003). Face recognition: Comparing component-based and global approaches. *Computer Vision and Image Understanding*, *91*(1/2), 6–21.

Kanade, T. (1973). *Computer recognition of human faces*. Basel and Stuttgart: Birkhauser.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA: MIT Press.

Leder, H., & Bruce, V. (1998). Local and relational aspects of face distinctiveness. *Quarterly Journal of Experimental Psychology*, *51A*, 443–473.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *Quarterly Journal of Experimental Psychology*, *53A*, 513–536.

Lee, D., & Seung, H. (1999). Learning the parts of objects by non-negative matrix factorization. *Nature*, *401*, 788–791.

Lobmaier, J. S., & Mast, F. W. (2007). Perception of novel faces: The parts have it! *Perception*, *36*(11), 1660–1673.

Lowe, D. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, *60*(2), 91–110.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, *6*(6), 255–260.

Phillips, P. J. (1999). Support vector machines applied to face recognition. *Advances in Neural Information Processing Systems*, *11*, 803–809.

Rakover, S. S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology*, *93*, 1–30.

Rhodes, G., Brake, S., & Atkinson, A.P. (1993). What's lost in inverted faces? *Cognition*, *47*(1), 25–57.

Rhodes, G., Tan, S., Brake, S., & Taylor, K. (1989). Expertise and configural coding in face recognition. *British Journal of Psychology*, *80*, 313–331.

Riesenhuber, M., Jarudi, I., Gilad, S., & Sinha, P. (2004). Face processing in humans is compatible with a simple shape-based model of vision. *Proceedings of the Royal Society London B (Suppl.)*, *271*, S448–S450.

Rossion, B., & Gauthier, I. (2002). How does the brain process upright and inverted faces? *Behavioral and Cognitive Neuroscience Reviews*, *1*, 63–75.

Schwaninger, A., Carbon, C. C., & Leder, H. (2003). Expert face processing: Specialisation and constraints. In G. Schwarzer & H. Leder (Eds.), *Development of face processing* (pp. 81–97). Göttingen: Hogrefe.

Schwaninger, A., Lobmaier, J. S., & Collishaw, S. M. (2002). Role of featural and configural information in familiar and unfamiliar face recognition. *Lecture Notes in Computer Science*, *2525*, 643–650.

Schwaninger, A., Wallraven, W., & Bülthoff, H. H. (2004). Computational modeling of face recognition based on psychophysical experiments. *Swiss Journal of Psychology*, *63*(3), 207–215.

Schwaninger, A., Wallraven, C., Cunningham, D. W., & Chiller-Glaus, S. (2006). Processing of identity and emotion in faces: A psychophysical, physiological and computational perspective. *Progress in Brain Research*, *156*, 321–343.

Searcy, J. H., & Bartlett, J. C. (1996). Inversion and processing of component and spatial-relational information of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(4), 904–915.

Sergent, J. (1984). An investigation into component and configurational processes underlying face recognition. *British Journal of Psychology*, *75*, 221–242.

Sergent, J. (1985). Influence of task and input factors on hemispheric involvement in face processing. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(6), 846–861.

Swets, D., & Weng, J. (1996). Using discriminant eigenfeatures for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *18*, 831–836.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology*, *79*, 471–491.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory and Cognition*, *25*, 583–589.

Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, *3*, 71–86.

Ullman, S., Vidal-Naquet, M., & Sali, E. (2002). Visual features of intermediate complexity and their use in classification. *Nature Neuroscience*, *5*(7), 682–687.

Wallraven, C., Schwaninger, A., & Bülthoff, H. H. (2004). Learning from humans: computational modeling of face recognition. *Proceedings of early cognitive vision workshop, ECVW 2004*, .

Wallraven, C., Schwaninger, A., & Bülthoff, H. H. (2005). Learning from humans: computational modeling of face recognition. *Network: Computation in Neural Systems*, *16*(4), 401–418.

Weyrauch, B., Heisele, B., Huang, J., & Blanz, V. (2004). Component-Based face recognition with 3D morphable models. *Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)*, *5*(85).

Wiskott, L., Fellous, J., Krüger, N., & v. d. Malsburg, C. (1997). Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern and Machine Intelligence*, *19*(7), 775–779.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception*, *14*(6), 737–746.