***Unfortunately, with the present article we have not yet crossed the Rubicon, but our hope is that we are no longer on the bank either***

**Bachmann (1989)**

# 5 General Discussion

*Early processing*

The first and most important conclusion from this series of experiments is that processing of faces is already possible by presenting faces for only some hundredths of a second. Experiment 1 showed that changes to the eyes are already detectable after a presentation time (PT) of 32 ms. Moreover, Experiment 3a revealed that the recognition of familiar faces is already possible after a PT of 26 ms. This demonstration supports other findings that have emphasized early processing of faces (e.g., Biederman, 1981) or natural scenes (e.g., Delorme, Richard et al., 1999).

*Microgenetic processing*

The second finding is that this fast processing does not occur at once, but as a microgenesis of recognition. It was revealed that the starting point for the processing of single facial features differs (Experiment 1). For instance, eyes (E) were found to be processed with higher priority than noses (N). A possible reason for this priority of features might be the differential information content or the diverse saliency of facial features. In fact, saliency of a feature had a strong influence on the recognition performance as well as on the recognition speed. Highly salient facial features were processed faster and more reliably than features of low salience.

*Configural vs. local processing*

Moreover, the nature of processing was dependent on the quality of facial information. Participants had to detect changes to faces in Experiment 1. These changes were achieved by *local* replacements of features or by *configural* alterations. Recognition was faster and more accurate when the faces had been changed locally. This might not be very intriguing, if the salience of configural and local changes had not been evaluated as being comparably high in pre-studies (Pre-Study 1a and Pre-Study 2). Thus, the differential effects of configural and local faces were *not* due simply to differing degrees of salience. Furthermore, highly salient configural faces had been evaluated as being less plausible and less attractive than their local faces of comparison. It is conceivable that highly implausible faces have a greater tendency to be conspicuous. This seems to be the case if faces are viewed as a set of faces with unlimited time resources. In the time-limited *change-detection task* of Experiment 1, however, this could not be shown. On the contrary, even highly salient configural faces were processed later and with lower accuracy than local faces. Thus, although configural information plays a key role in face recognition (cf. section 3.1.3), it seems that this type of information is not processable very deeply within the first 100 ms.

*Parallel vs. serial processing*

Moreover, features of a configurally changed face seem to be processed simultaneously, whereas those of a 'local face' were found to be processed serially. It is important to note that only four different processing models were tested and compared. Of course, alternative models are conceivable, especially with chimerical processing assumptions (see section 3.2.3), which were not addressed here. Furthermore, as Wolfe (1996) has demonstrated, the pattern of results produced by a serial self-terminating search can also be produced by a variety of limited-capacity parallel models (see also Kinchla, 1974; Ratcliff, 1978; Ward & McClelland, 1989). However, in the following discussion we have to focus on the findings concerning the explicitly tested processing models.

The parallel process (for the configurally changed faces) seems to start with processing of the eyes first, later followed by the processing of the mouth and nose—while the processing of the eyes continues. Despite this finding, locally changed faces seem to be processed in a strictly serial way. Again, the eyes were recognized first. After the eyes had been processed, the mouth was processed. Finally, the nose was recognized. Therefore, this order will be called E→M→N. Such a scanning strategy is rather complex, because it 'jumps' from the eyes—which are the highest internal feature, regarding spatial position—to the lowest (mouth) and after that to the central area (nose) of the face. A much simpler strategy would be to scan linearly from the top of the face to the bottom (e.g., Smith & Nielsen, 1970; Rhodes, 1988). Besides, the position of the least recognized area, the nose, was situated exactly in the center of the picture. Therefore, an alternative simple strategy for the processing of a face would call for the recognition process to begin in the center of the stimulus. However, this was not found to be the case.

It is questionable whether the assumed processing strategies are hard-wired or whether they can be adapted by exceptional stimuli or cognitive penetration. A hint for deviations from the sequence of E→M→N was found for one locally changed face. Face M03 was equipped in the high-salience variant, due to a construction error, with a mouth made up with a bright red color. For this face, an alternative serial strategy was discovered. Again, the processing was strictly serial, but it now followed the sequence M→E→N. This is an indication that the processing sequence is *not* hard-wired. Moreover, it seems that an ordinary sequence of E→M→N is most useful for normal and not very deviant faces. Nevertheless, an adaptation of this sequence in favor of extraordinarily high degrees of salience is possible—and seems to be functional. Whenever a face contains unexpected and highly salient features or areas, it engenders a deviation from the normal path of processing. Therefore, such features are *a priori* informative and important. Indeed, this is the true reason for coloring the skin, especially accentuating the mouth in a distinctive way: to attract attention!

*Alternative process assumptions*

This last argument for an adaptive processing strategy challenges an alternative-processing model. This model is also capable of explaining the processing sequences, at least for the prioritizing of eyes features, but it is much less informative than the model outlined above. One might argue that the assumed processing sequences are *only* caused by artifacts of salience and nothing else. Not the features as such nor their quality are essential factors, but the pure *salience* of them. A more highly salient feature will be processed earlier and more efficiently than a less salient one. This hypothesis seems to be rather intuitive, but it is nevertheless not fully compatible with the empirical data. Salience is indeed an important factor in the recognition process, as the recognition data between high- and low-salience features have demonstrated (Experiment 1 and Experiment 2). Nonetheless, the overall salience of faces in which the eyes had been changed was not found to be different from those faces in which the mouth had been changed (Pre-Study 2). The recognition process between both faces differed, however, as revealed by the recognition data in Experiment 1. Furthermore, the plausibility and attractiveness were much lower for highly salient *configural* faces than they were for highly salient *local* faces. It appears reasonable that a very implausible or very unattractive face would stand out and would therefore constitute a very salient cue. The opposite was the case in regard to speed of processing and recognition performance! Configural processing started later and proved to be on a lower performance level. Therefore, it has to be assumed that the processing of configural and local features is indeed based on a different quality or nature.

*Local feature analysis*

The second experimental series using Thatcher faces revealed that early local feature analyses were beneficial for a fast and accurate processing of faces. This is in accord with findings of Cooper and Biederman (1993): in a task with limited presentation times ($t$=100 ms), the subjects had to judge whether two images had the same or a different name. It was shown that featural changing had a greater salience over metric properties (see also Biederman et al., 1999).

In the present work, the early local feature analysis was tested with *inverted* face stimuli. This seems rather unrealistic for investigating face recognition *per se*. In everyday life, faces are commonly oriented upright. Thus, are these results transferable to face recognition in general? Valentine and Bruce (1988) found a linear increase in the RT of same-different judgment as the second of a pair of sequentially presented faces was rotated away from the vertical. Moreover, it was found that the effect of changing facial expression did *not* interact with orientation. These results were interpreted as evidence against the view that inverted faces are processed in a qualitatively different manner from upright faces as proposed by Carey and Diamond (1977). The results of the Valentine and Bruce (1988) study are more consistent with the view that recognition of upright and inverted faces depends upon the processing of the same features (cf. Goldstein & Chance, 1980; Flin, 1985). Nevertheless, it must not be forgotten that other studies have demonstrated a reduction of configural processing in inverted faces (Leder & Bruce, 2000b). Recently, the size of the face inversion effect was even *operationalized* as a measure for involvement of configural processing (Mondloch et al., 2002). Therefore, it might be argued that local feature analyses would be *a priori* advantageous in the processing of inverted faces, because configural processing components were hardly available, thus recognition had to be based on the quality of the remaining information. This might be the case to some degree. However, the initial question was not whether local feature analyses are *available* in early face processing, but whether they are *beneficial* for the further face recognition. The answer to this question is: Yes, they are! In Experiment 3a, Thatcher faces were recognizable faster than original faces. Furthermore, Experiment 5 demonstrated that this RT advantage is so large that it can compensate for the global precedence that is usually discovered (e.g., Parker & Dutch, 1987). Thus, local feature analysis is indeed an important component of early face processing!

### *Holistic processing*

When the PT was extended to 200 ms, the local feature advantage vanished. For Experiment 3a it was argued that under these circumstances holistic processing would emerge. Local feature analyses might still be present, but then the template-like recognition mode of original faces is even more advantageous. This concept of holistic processing differs slightly from the original formulation by Tanaka and Farah (1993) and Farah et al. (1995). In the original view, holistic processing just meant that parts are not represented explicitly. In contrast, holistic processing according to the results of Tanaka and Sengco (1997) and Rhodes et al. (1993) would imply that component and configural information are first encoded separately and then integrated into a holistic representation. This is in line with the results found by the experimental series with familiar faces in the present work. In a first stage, global and local information are processed separately and are later combined into a holistic face (recognition). It is important to stress this point. Holistic recognition is not a one-stage processing mode, but it seems to be the combination of many preceding processing stages. The supposition that binding of several information qualities is established relatively late is not new. Marr's *theory of vision* (Marr, 1982) or the *cascade processing model* of Humphreys et al. (1988) follow similar ideas. Interestingly, Humphreys and Riddoch (1987) have also demonstrated that clinical cases of visual object agnosia were unable to combine visual information to complete objects (cf. Riddoch & Humphreys, 1987). For these cases, the essential step for binding several information qualities was apparently not achievable *at all*. Furthermore, Brooks et al. (1997) have assumed that featural as well as holistic types of processing are specialized for different face recognition strategies, each. According to Brooks et al., holistic processing is well suited for face *matching* tasks, whereas featural/local processing is optimal for recognizing face *details*.

### *Outlook*

The results of the experiments presented here provide new evidence for different face processing strategies and their differential temporal importance. Nevertheless, this can only be seen as the beginning of exploring early processing of face recognition. However, the experimental designs used here were rather restricted in terms of their tasks and the materials, in order to be able to test specific hypotheses in a systematic manner.

One of these constraints was the use of normalized, frontal, and static faces. It is questionable whether the stimuli used here were processed as pictorial or structural codes. The distinction between pictorial and structural codes has been made by Bruce (1982) and by Bruce and Young (1986), who argued that a *pictorial code* is an episodic representation of a particular visual stimulus (such as a photograph), whereas a *structural code* captures more elaborate aspects of the stimulus. In terms of a facial stimulus, these aspects would constitute higher-level information, which is independent of changes in viewing angle, lighting, orientation, and so on. Although the properties of pictorial codes are certainly of interest, for example to systematically test face processing hypotheses, Bruce and Young (1986) argued that they are unlikely to contribute to everyday face processing. In everyday life, the exact conditions under which a particular face was previously encountered will seldom be repeated, nor will a face recognition task be restricted to a very limited set of faces. Young et al. (1985) have argued that simultaneous matching tasks are particularly suspicious because of their reliance on pictorial codes. The tasks used in the previous empirical studies were not of this special type of simultaneous matching. However, it remains questionable whether the strategies used in these tasks resemble more closely pictorial coding or structural coding. Nevertheless, due to the sequential presentation mode in combination with visual masking, short-term memory must have been involved. Therefore, the processing seems to not be operating on the pictorial level only.

Moreover, for investigating the processing sequence of locally changed faces more deeply, it would be advisable to use a more naturalistic operationalization of saliency. One idea to improve the naturalness would be to use affective facial features instead of generating salience through purely physical factors, such as coloring the features. An important subsequent question would then be: are such features also capable of reversing the processing sequence outlined above?

Furthermore, the static and frontal appearance of faces in the experiments does not seem to be of high ecological validity. In everyday life, such conditions are rarely attainable. We live in a dynamic 3D world with enormous everyday recognition problems. Our everyday task is to respond to these requirements quickly, without much effort, and with high accuracy. This is essential in order to be able to live in a social community and, last but not least—to survive!

Recent investigations in this direction have resulted in studies of facial motion (e.g., Lander & Bruce, 2001; Lander et al., 1999), the scrutiny of 3D-face effects (e.g., Bruce et al., 1987; Leder & Carbon, subm.-b), or the testing effects of facial expressions (e.g., Ogawa & Suzuki, 1999; Calder et al., 2000).

It is still a long and winding road, however, to reaching a consummate understanding of face recognition—a capability that is acquired naturally, yet proves to be highly complex when scrutinized scientifically.

Despite all these restrictions, in closing I agree with William Uttal:

> "A review of the literature leads to the conclusion that I still have not answered some of the most basic questions involved in how we see faces. The author [W. Uttal] concludes, furthermore, that truly reductive theories of face perception are not possible and that most contemporary theories of face perception are, at best, behavioral descriptions. It is concluded that face perception is a multidetermined process to which both local and global attributes may contribute" (Uttal, 2001).

The present work has presented different new paradigms and results in order to investigate early face processing in more detail. Therefore, I hope that they provide at least a small step towards a better understanding of this field.