4 Empirical Studies

In the following part of this work, several experiments will be presented which investigate the early processing in face recognition. The structure is as follows: first, open questions in the field of face processing will be discussed (chapter 4.1). In chapter 4.2, a short introduction to general methodical aspects of this work is given. After these more general parts, specific experiments are described that investigate the open questions. This will be done in two separate parts. One investigates microgenetic face processing models with unfamiliar faces (chapter 4.3); the other examines hypotheses dealing with local face analyses of famous faces (chapter 4.4).

For the experiments using unfamiliar faces, artificial facial material was constructed. In order to create standardized facial material, the distinctiveness of specific facial parts was evaluated and combined systematically (Pre-Study 1a and Pre-Study 2). A *change detection* paradigm with varied presentation times (PTs) was used to investigate the role of configural vs. featural processing within the first 100 ms in order to construct a microgenetic model of face processing (Exp. 1 and Exp. 2). In a second series of experiments (Exp. 3a, Exp. 3b, Exp. 4, and Exp. 5), the specific role of local features in the early face processing was investigated. Furthermore, the contextual and configural embedding of the local features were analyzed. To test these hypotheses, a *Thatcher face* paradigm combined with variation in presentation times (PTs) was used.

Most importantly, in all experiments a systematic variation of features and configuration in combination with the usage of limited presentation made it possible to test specific processing hypotheses.

4.1 Open questions

In the previous chapters, theoretical and empirical accounts of face recognition were introduced. There are many open questions left in this field of research, especially in respect to the *processing* of faces. The aim of this section is to briefly address some of these unresolved problems. Moreover, it will be explained how the following studies and experiments attempt to investigate these problems.

How much is processed in early processing?

In section 2.3, the early processing of objects and faces was discussed. It was demonstrated that within only 100 ms of presentation, simple structures were already recognizable. Biederman (1981) has even demonstrated that a presentation time of about 80 ms (20-80 ms according to Delorme, Richard et al., 1999) is generally sufficient to recognize general structures and semantics of a more complex natural scene. In this work, this time period of the first 100 milliseconds of a presentation will be investigated further, particularly in respect to *what* kind of information is available.

Role of distinctiveness in face processing

Distinctiveness seems to be an important factor for the nature of ongoing processing (see section 3.1.1). In Experiment 1 and Experiment 2, the impact of the distinctiveness of several features on the face recognition process will be studied.

Microgenesis

In recent years, the microgenetic approach to pattern recognition has become quite popular (Sergent, 1986b; Watt, 1988), as described in section 3.2. The main idea behind this approach is that different levels or aspects of the image become perceptually available at different moments of real time while this accumulative process of percept development is occurring (Bachmann, 1991). Thus it is assumed that our visual representation of a scene is not achieved in one step. It is attained incrementally. One of the most important issues of the present work will be to analyze the microgenesis of early face recognition. Therefore, in Experiment 1 concrete processing models will be tested. The logic behind these models is a microgenesis of the processing of single face features or face regions.

Relationship between configural and local information

As described in section 3.1, there are some hints that configural/relational and featural/local aspects of a face are dissociable and are *not based* on the same cognitive processes. The question is whether configurally or locally changed faces also differ in terms of the nature of their processing. Experiment 1 and Experiment 2 will investigate this question with systematically constructed faces, which are manipulated in either configural or local aspects.

Moreover, in supplementary experiments, the special role of local feature analysis (see section 3.1.1) and its relationship to configural information will be analyzed further. One prominent question is whether both information qualities are already binded if faces are only presented very briefly (see section 3.2.3).

Global precedence

Section 3.2.3 described many processing examples in which global structures preceded local structures. In Experiment 4 and Experiment 5, the involvement and quality of global processing will be tested. Moreover, the relationship between global processing within different presentation time constraints will be studied.

Holistic processing

The assumption of holistic face processing is intuitive and has often been validated. Nevertheless, the important question remains whether holistic processing occurs even after only a short glance (see section 3.1.3). The experimental series using familiar faces will search for beneficial holistic processing in comparison to specialized local feature analyses.

4.2 Methodological Introduction

The central issue of this section will be to discuss methodological aspects of the experiments described and discussed in the empirical part of this work. Its aim is to explain why and under which specific conditions particular stimuli will be used. Furthermore, it has to be discussed which general methodical problems are expected to emerge in such experiments, and how they can be solved.

4.2.1 Stimuli

The muscles in our face enable us to make up to 7000 distinct expressions (Ekman et al., 1972). But just as is the case for our spoken vocabulary, in normal day-to-day life we only use a small proportion of this repertoire—a few hundred expressions (Bates & Cleese, 2001). Expression analysis and the impact of expression on face recognition is an interesting but also vast field of research (Frijda, 1986).

In the present work, solely the early processing in the recognition of faces will be investigated. To do so, neutral and static faces will be used that were photographed frontally. These faces will be highly standardized. Many results have shown that transformed faces with changed expression are identified less accurately than untransformed faces (Parkin & Goodwin, 1983). These factors, despite their advantage of improving the *ecological validity*, decrease the inner validity and the possibility to obtain recognition rates above the base rate when a task is very difficult. Static faces were employed as well (as argued in Knight & Johnston, 1996) to focus on the small areas of interest, the eyes, the nose, and the mouth, whereas movement studies are important to study the more elaborate face recognition performance after early processing is completed (e.g., Lander, Christie, & Bruce, 1999; Lander & Bruce, 2001). The focus can be obtained more easily when there is great homogeneity in the stimuli.

Moreover, to make the pictures of the faces as realistic as possible (regardless of the constraints discussed above), only photographic materials were used in the present studies. Bachmann (1991) pointed out that *Identikit* faces as well as other frequently used schematic faces might be processed in a different way than natural faces. The use of schematic faces poses some serious problems in regard to the generalizability of the conclusions drawn from these studies. First, it is well known that texture (Hill & Bruce, 1996) and color (Lee & Perrett, 1997) play an important role in face recognition, and it is desirable to use realistic stimuli. Under some conditions, impoverished representation leads to an unrealistic, differential weighting of facial features (Leder, 1996). In addition, all faces in the present work were Caucasian, to avoid racial stereotyping (Zebrowitz et al., 1993) and to refer to the same race expertise (Valentine, 1991; Valentine et al., 1995; Valentine & Bruce, 1986b; O'Toole, Deffenbacher, Valentin, & Abdi, 1994; Levin, 1996; Hill et al., 1995; Dehon & Brédart, 2001; Bothwell, Brigham, & Malpass, 1989). See section 2.2.1 for more details.

Unfamiliar vs. familiar faces

Bruce and Young (1986) support the view that different functional components are responsible for the recognition of familiar faces than for the temporary storage and matching of representations of unfamiliar faces. Indeed, familiar and unfamiliar faces could be dissociated neurologically (Paller, Gonsalves, Grabowecky, Bozic, & Yamada, 2000; Young, Hay, McWeeny, Flude, & Ellis, 1985). In a very early finding, Warrington and James (1967) showed that although lesions of the right cerebral hemisphere were able to affect the recognition of familiar and unfamiliar faces, such impairments of familiar or unfamiliar face recognition did not correlate with each other. Familiar faces are generally found to be recognized with a right-hemisphere advantage in contrast to unfamiliar faces (cf. Carey & Diamond, 1994; Moscovitch, Scullion, & Christie, 1976). Other results validate the dissociation between both kinds of faces. For example, recognition of unfamiliar faces is clearly dependent upon viewpoint (Troje & Bülthoff, 1996; Hill & Bruce, 1996; Hill, Schyns, & Shigeru, 1997; O'Toole et al., 1995), even if some pictures only differ with respect to illumination (Troje & Bülthoff, 1998). Exceptions to these findings were found by the Biederman group (e.g., Biederman & Gerhardstein, 1993) and by Bruce, Valentine, and Baddeley (1987).

It would seem that people are able to form a more effective representation of the structural properties of familiar faces (Young et al., 1985). The explanation for this is simple. Having a preexistent representation in memory is important not because it is available to be reactivated repeatedly, but rather because its existence allows for deeper and more elaborate encoding of the item at its first presentation. A familiar and meaningful item leaves a stronger and more accessible memory trace than a totally unfamiliar one (Bentin & Moscovitch, 1988). Therefore, familiar faces are better recognized than unfamiliar faces (Klatzky & Forrest, 1984).

As a consequence, for all the *identification tasks* in the present work, only very familiar faces were used for which a fast and reliable processing is expected. For *feature interchange paradigm* tasks however, where any familiar associations are unwelcome and the pure visual process is to be analyzed, unfamiliar faces were used.

Masking

In all of the following time-relevant experiments, the targets were presented in a clear-cut temporal range. The main paradigm that is helpful in this regard is the *method of limited presentation* (Turvey, 1973). As Haber (1969) pointed out, it is the unique technique whereby one can supposedly halt stimulus processing at a certain stage without relying on the subject's ability to describe what has already been seen. To implement such a limitation, a type of mask was used, as described in Loftus, Duncan & Gehrig (1992), that appeared abruptly after the stimulus introduction. This was done to overcome the problem of incalculable 'fading out' persistence of the image stored in iconic memory (Loftus, Johnson, & Shimamura, 1985). For example, Coltheart (1980) has estimated the duration of the iconic trace at about 300 ms, which is rather long and would be problematic when investigating very brief presentation in about 10 ms steps as in one of the experiments below (cf. Baddeley, 1998).

Moreover, the masking technique is useful to prevent a pure change detection strategy. It is well known that the visual system is very sensitive to events that exhibit sudden changes (Theeuwes et al., 1998; Breitmeyer & Ganz, 1976). However, with an interstimulus mask, participants have to rely to a larger extent on mental representation. With visual masking, a qualitative step is possible from perception to memory, which is a prerequisite for the experiments presented in this work. Aside from these advantages, however, it must not be forgotten that a visual mask can also function as an interference stimulus, having the power to disrupt or bias the processing of the actual stimulus (Eriksen, 1980). Marcel (1983) speculated about an impeding of perceptual processing through visual masking (see also Nisbett & Wilson, 1977), and Julesz (1981a) advised to refrain from using such a technique.

Brief presentation

Raymond Dodge (1907) has shown long ago that information acquired during each fixation should have time to be 'cleared up' in subjects' conscious perceptual image. The main motive for working with a brief physical exposure time, as discussed above, is to reduce the stimulus and the consequent psychophysical process to the simplest possible terms. The experimenter seeks to isolate a single cognitive or apperceptive event to eliminate, if possible, all biasing changes in the position of the visual point of regard, and in the direction of attention.

In the prophetic novel *1984,* George Orwell foretold of a future in which our thoughts, attitudes, and behaviors would be controlled almost entirely by the media. This startling prophecy seemed to gain plausibility in the late 1950's after James Vicary reported significant increases in Coca-Cola sales with the help of his subliminal advertisement method of flashing 'Drink Coke!' directives during movies and advertisements (see Packard, 1957). Although the striking success of this technique could never be replicated convincingly, it seems that we do indeed process much information subconsciously or, technically speaking, subliminally. As early as 1884, Pierce and Jastrow (1884) reported that people could perceive small differences in pressure to the skin without conscious awareness of differing sensations. Most importantly, many experiments have demonstrated that it is possible, in some cases, that a person uses the current visual input to produce a relevant motor output, without being able to say what he or she has seen (Crick & Koch, 1998).

Sperling (1960) developed an experimental procedure to address the dilemma of determining what people had seen during a brief presentation. It is a real dilemma, Sperling pointed out, because the question 'What did you see?' requires the observer to report both what he or she remembers and what he or she has forgotten (Sperling, 1960, p.1). Sperling very briefly (for 5-500 ms) presented to his subjects an array of letters arranged in a matrix. Two different types of recall instructions were used. In full-report, subjects were simply told to report as many letters of the entire array as they could remember. Sperling found that subjects could usually recall an average of about 4.5 items, regardless of display size or stimulus duration. He referred to this quantity as the *immediate memory-span* (Sperling, 1960). In the partial-report, the participants were shown a randomly-cued subset and were asked to recall numbers from it. By contrasting these methods, Sperling found that subjects could usually recall about 75-90% of the items when cued. He concluded that full-reports might be invalid to test the actual memory performance. Therefore it is critical to make recall or memory tasks as easy and uncostly as possible to prevent biasing memory decays. Sperling (1960) additionally pointed to the pre-categorical nature of information stored in the so-called *iconic memory* (Neisser, 1967), which is very basic and physical (see for other explanations Coltheart, 1980). Pylyshyn (in prep.) agreed with the assumption of a rudimentary quality of the given stimuli during a brief presentation, but suggested that they be not only pictorial, but also developed from geometrical incompleteness and imperfection to higher degrees of visual richness through longer presentation times (cf. section 3.2 about microgenesis).

How long does it take to create a percept?

Biederman (1981) has demonstrated in a number of experiments that a glance of about 80 ms (20-80 ms according to Delorme, Richard et al., 1999) is usually sufficient to recognize general structures and semantics of a natural scene. Since we have seen that a saccade needs at least 200 ms to be initiated (see Wolfe, 1996; Rayner, 1983), only a single fixation of the eye is required to capture the scene. Of course, the exact presentation time needed to recognize a stimulus varies according to different stimulus classes, quality of the pictures and their distinctiveness. Finally yet importantly, the presence or absence of a subsequent mask and the type of mask serves to bias the psychologically relevant presentation time (cf. paragraph about visual masking above).

Locher (1993) transferred brief presentation evidence for identification processes to the field of aesthetics. He found that students were able to differentiate between levels of attractiveness on the basis of cue information contained in a 'single brief glance' of 100 ms. This lends support to the idea that already within such brief PTs higher cognitive processes may occur. In accord with the microgenetic account, however, it seems that within such a short time the full information of a scene or an object is not yet available; the processing remains on an imperfect and deficient status of recognition. For example, words presented very quickly (33 ms) and masked afterwards are analyzed mainly at the level of word *parts*, not *whole* word meaning (Abrams & Greenwald, 2000). Consequently, it is plausible that subliminal primes receive analyses that operate on no more than parts of words or objects, respectively (see also Abrams, Klinger, & Greenwald, 2002).

To systematically investigate how much is available within a certain time period, the *method of limits* is useful. In this method, the stimulus intensity, or in this particular case the *presentation time*, is either increased or decreased in small increments. On the ascending trial, the intensity is set below the subject's threshold and is increased by small amounts until the subject reports sensing the stimulus (Snodgrass, Levy-Berger, & Haydon, 1985). This method was derived from the methodical apparatus of Fechner (1860) and later refined by Jastrow (1888).

The quality of briefly presented stimuli

As is obvious from the sections above, briefly presented stimuli seem to be incomplete and not fully processed at once, but elaborated in a microgenetic account. It has been shown, for instance, that the earliest information arriving in each cortical area is not processed in terms of color (Delorme, Richard et al., 1999; Delorme, Richard, & Fabre-Thorpe, 2000a). Removing the color information from natural images had only mild effects on rapid categorization in both monkeys and humans (Delorme, Richard, & Fabre-Thorpe, 2000b). Treisman et al. (1992) was also able to demonstrate that, above all, preattentive processing is a bottom-up process which reflects the activation of populations of feature detectors. This means that features might be the earliest available structures in brief presentation (cf. discussion about global-to-local in section 3.2.3).

4.2.2 Dependent variables

4.2.2.1 Accuracy

In the following section, the accuracy measures used in this work will be discussed briefly.

Percentage correct rate

The main accuracy variable used in the present work is *percentage correct*, a measure that aggregates *hits* as well as *correct rejections*. Since there is no hint on how much of the amount of correct answers is made by guessing, a second measure is used in conjunction with the correctness rate. The rate of false alarms gives us insights into the quality of reasoning. Dailey, Cottrell and Busey (1998) assumed that the analysis of false alarms or errors is essential for our understanding of the representations and mechanisms underlying face processing. Moreover, to this day it is not very clear what ERP or FFA-fMRI data stands for and which cognitive processes and performance they reflect (Kanwisher, Tong, & Nakayama, 1998). As a consequence, a well-established measure like percentage correct is more intuitive and suitable for the purpose of the present work.

Sensitivity measurement

There are many ways of combining measures of correctness and falseness in one single variable (Norman, 1964). The advantage of such a measure is that it is easier to work with, albeit at the cost of thus obtaining a more abstract nature.
The two most commonly used measures are *d'* (called *d-prime*) and *A'* (*A-prime*). *d'* is given by Equation 4-1.

Equation 4-1: Calculating *d'*.

$$d' = z(H) - z(FA) \qquad ;$$[31]

This equation is known as the *Equal-Variance model* for calculating *d'* (Simpson & Fitter, 1973). This model is in accord with the Gaussian signal detection theory (SDT) and contrary to the account of a *two-threshold* theory (Luce, 1963) or *three-threshold* theory (Krantz, 1969), where no sensory threshold is postulated (see Swets, 1961; Tanner & Swets, 1954).
SDT uses the hit rate and the false alarm rate to estimate a second measure. Besides the measure of discriminability *d'*, there is the measure of bias, which is called *C* (or, in an older ter-

---

[31] *H* stands for Hits, *FA* for False Alarms.[32] There is a convenient Java applet available in the WWW that uses a method suggested by Snodgrass, Levy-Berger, and Haydon (1985) and Brophy (1986) to calculate *A'* (as well as *d'* and *C*) and a method suggested by Donaldson (1992) to calculate *B''D*. See http:// rumpole.psych.purdue.edu/models/RecognitionMemory.html for an online-calculation.

minology, β). The larger *d'* is, the better is the subject's ability to truly discriminate between old and new items. Values of *C* above zero indicate a conservative bias, which implies a lower willingness to guess 'old', whereas values of *C* below zero indicate a liberal bias. As an exception, *d'* cannot be calculated when there are hit or false alarm rates of 1 or 0.

This limitation is not valid for an alternative discriminability measure, called *A'* (see Snodgrass & Corwin, 1988; Weber, 1993; Donaldson, 1992; Rae, 1976; Valentine & Endo, 1992). In addition, *A'* does not require homogeneous variance like *d'* does. *A'* varies from 0 to 1 with 0.5 indicating chance performance. The corresponding measure of bias is called B''_D (or *Pr* , see Feenan & Snodgrass, 1990).

In the present work, the *A'* measure for evaluating discriminability will be used according to the formula presented in Leder and Bruce (2000a, p.162)[32]. See the formula for calculating *A'* in Equation 4-2.

Equation 4-2: Calculating *A'*.

$$A' = \frac{H^2 + FA^2 + 3 \cdot H - FA - 4 \cdot FA \cdot H}{4 \cdot H \cdot (1 - FA)}; \qquad \text{if } H \geq FA;$$

$$A' = \frac{H - H^2 + FA - FA^2}{4 \cdot FA \cdot (1 - H)}; \qquad \text{if } H < FA;$$

4.2.2.2 Reaction Time (RT)

> ***Would thought also not have the infinite speed usually associated with it, and would it not be possible to determine the time required for shaping a concept or expressing one's will? For years this question has intrigued me.***
>
> **F.C. Donders (1969)[33]**

With this comment, F.C. Donders introduced the field of mental chronometry to psychology in a paper published in 1868 entitled *On the speed of mental processes* (Snodgrass et al., 1985). He used the so-called *subtraction method*, which has been replaced later by Saul Sternberg's (1969) *additive factors method* (see also Snodgrass et al., 1985).

There are some difficulties in the reaction time[34] (RT) data collection that have to be discussed briefly, because RT is one of the essential dependent variables measured in the present work.

RT measurement

Using reaction times as a dependent variable precludes the separation of the respective durations of perception, decision, and motor response stages (Luce, 1986). Because we cannot usually set limits to the complexity of the mental processes in question, one major strategy is to measure the speed of mental processing by isolating some processes. These processes must meet two vital criteria. First, they have to be sufficiently elementary, and second, they have to

---

[33] Originally published in 1868.

[34] I will use the term *reaction time* (RT) instead of *response time*, as does Luce (1986, p.2), in referring to response times collected in experiments. Other authors have proposed the term *inspection time*, which is the minimum time required to make a single observation or inspection of the sensory information with an accuracy of 97.5% (Vickers, Nettelbeck, & Willson, 1972), but I will use pure RTs only.

be relatively immune to the influence of higher cognitive activities or of motivational and social factors (Vickers & Smith, 1986).

Macho and Leder (1998) thought of RTs as one of the most simple assumptions concerning the relationship between theoretical entities and behavior, because theoretical entities are directly reflected by RTs. Moreover, RT measurement is a highly intuitive variable, which does not need such a complex apparatus as, for instance, *event-related lateralisations* in the EEG, which is also a precise measure of attentional processes independent of RTs (Wolber & Wascher, 2002).

To improve the data quality and to decrease the noise, warning signals like fixation crosses are often used to announce that the experimental stimulus is about to emerge (already proposed by James, 1905). Another important remedy against noise in data is to calculate RT for correct trials only (e.g. Watson & Humphreys, 1997). It is likely that *same trials* and *different trials* differ in respect to their underlying processes. Therefore, they seem to be distinct from each other. The same seems to be the case with *correct trials* and *false trials*, where different strategies and cognitive processes might be involved. Therefore, in the present work, primarily RT data of correct same trials are analyzed.

Additionally, it is essential how the RTs are collected and analyzed. Most problematic are outliers in RT data. An extremely long RT, for instance, which is accomplished by a non-alerted or distracted participant, would change the average substantially. Therefore, the median is often used as centrality parameter instead of using the arithmetic means. Another strategy to cure this problem would be to set an outlier criterion. Such a criterion is used to limit the RT-range. On the one hand, a *static* criterion restricts RTs within a static RT-range. RTs could, for example, be restricted to lie within the range of 500 ms up to 5000 ms. On the other hand, *dynamic* criterions set the RT limit for each subject individually. For example, RTs have to lie within a symmetric range of ±*d SD*s (=standard deviations), where *d* is typically 2.5 *SD*s, 3 *SD*s or 4 *SD*s away from the averaged RT of the individual subject (see Snodgrass et al., 1985). The criterion of outliers is not standardized like the α-level. Thus, it must be adjusted to the given circumstances, and is not as strict and inflexible as an absolute limit (e.g., Rabbitt, 1966). If the pattern of the limited RT data is comparable to the pattern of the raw data (without an outlier criterion), it is known as ‘stable data’ that is not prone to participants’ response noise.

Simon Effect and right-left asymmetries

Spatial compatibility between stimulus and response is a very important determinant of performance. This can easily be demonstrated when subjects have to respond with a left or right key in response to a left or right stimulus. The speed and accuracy of their responses will then be in a close relationship between the position of the stimulus and the spatial compatibility with the pressed key, with compatible reaction being much quicker and more accurate than incompatible ones. Similar advantages of spatial stimulus-response correspondence can be observed when the responses are not related to non-spatial stimulus characteristics, for example color, shape, or attractiveness, in addition to a relatively unimportant position of the stimulus (e.g., Lu & Proctor, 1995). This effect is commonly known as the *Simon Effect* (Craft & Simon, 1970; for an overview see Lu & Proctor, 1995). Spatial correspondence can take on many forms, if only the stimuli and the responses correspond to their relative positions. Results suggest that visual stimulus selection and manual response selection are distinct mechanisms that operate on common representations (Hommel & Schneider, in press) in such a way that perceptual judgments even activate motor-response codes (Kerzel, Hommel, & Bekkering, 2001). It seems that motor responses and judgments are based on a common, cognitively penetrable, spatial representation.

Therefore, for all experiments of the present work, the response keys were balanced.

4.3 Experiments with unfamiliar faces

The experimental series with unfamiliar faces investigated the early processing of faces. There were two areas of focus. First, the relationship of processing between local features and configurational aspects of a face. Second, the processing sequence of features. To be able to test specific processing hypotheses, systematic material had to be constructed. In Pre-Study 1a, natural faces were evaluated in terms of their distinctiveness[35]. In Pre-Study 1b, the stimuli of Pre-Study 1a were studied with regard to material artifacts. For Pre-Study 2, new artificial faces were generated. This was done by systematically arranging material from Pre-Study 1a. Furthermore, it was tested whether the constructed faces fulfill the requirements for the latter experiments. Experiment 1 investigated the microgenesis and dependence of several facial features, whereas Experiment 2 extended the findings of Exp.1 by using longer PTs as in Exp.1.

4.3.1 Pre-Study 1a: Natural Faces

Introduction.

Distinctiveness of features

This study was designed to evaluate unfamiliar facial stimuli to obtain distinctiveness and other measures as well as attractiveness of this material. The participants had to rate 22 female facial pictures of the *DADA-Faces* database (Carbon, 2001) in respect to their overall and part distinctiveness as a psychometric measure (Guttman, 1971). Based on distinctiveness data, new artificial faces with parts of high, low, and average distinctiveness (eyes, nose, and mouth) were created for later experiments. By knowing the size of rated distinctiveness, it is possible to create standardized faces in terms of distinctiveness. The distinctiveness of a face and its parts has direct influence on the recognition rate in later tests (Bruce, Burton, & Dench, 1994). Therefore, it is important to know the distribution of distinctiveness for testing recognition and processing hypotheses. In several studies, particularly distinctive areas of a frontal face have been identified. The most distinctive parts seem to be the eyes (e.g., Brunelli & Poggio, 1993) and the mouth region (e.g., Scassellati, 1998). To test this for DADA-Faces, full faces as well as facial parts (eyes, nose, mouth, hair) were evaluated. A seven-point rating scale of distinctiveness was used as the dependent variable. This was done to get a pool of facial parts of high and low distinctiveness. Based on these groups, faces of low and high distinctiveness can be constructed.

Predictors for Distinctiveness

The following experiments will analyze the impact of distinctiveness on the face recognition process. Therefore, it is not only important to control distinctiveness, but also to identify valid predictors *for* distinctiveness itself. For that reason, distinctiveness ratings for facial parts are analyzed in terms of predictability of the overall distinctiveness. Additionally, the attractiveness of full faces was evaluated. Attractiveness is a complex and important psychological construct. Some researchers (O'Toole et al., 1998) have assumed that higher order semantic measures such as attractiveness might be a better predictor for distinctiveness than just simple physical measures.

---

[35] In the following work, the term *distinctiveness* will be used as a (dependent) variable; in contrast, *saliency* will be used as (independent) variable and as a factor of *distinctiveness.*

Face space

Another valid predictor for distinctiveness seems to be the inter-similarity between single faces. According to Valentine's (1991) hypothesis, a psychological face space should be mainly constructed in reference to the distinctiveness of faces: More distinctive faces are located in the periphery and typical faces in the centered area (see for an illustration Figure 3-1 in section 3.1). This hypothesis was tested with the similarity ratings of all possible face pairs.

Hotspots

Moreover, in this study a new distinctiveness measure (so-called *Hotspot*-procedure) will be introduced and evaluated. *Hotspots* indicate the most distinctive points of a face and are thus a very intuitive measure to illustrate distinctive regions. The *Hotspot* measure will be compared with the very common saliency measure of rated distinctiveness.

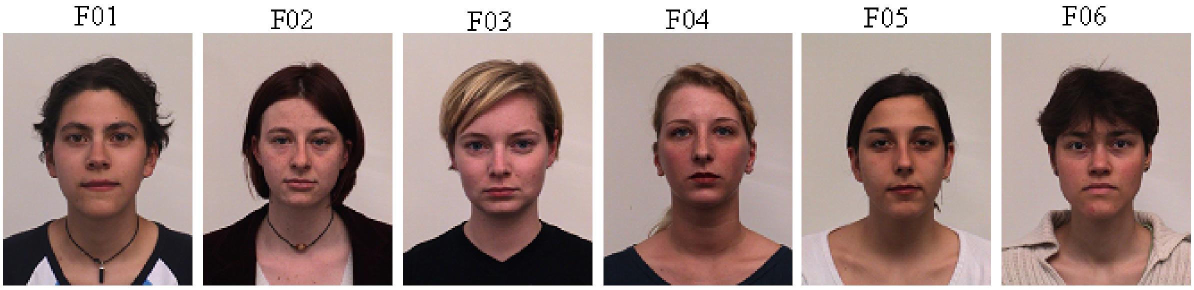
Method.

Participants. Sixteen students participated in the study. In all experiments, the participants were undergraduate students (12 women, 4 men) of the Freie Universität Berlin, who were given a course credit to fulfill course requirements. The mean age was 23.8 years (from 20 to 36 years). All participants had normal vision abilities or were corrected to normal vision. The faces used were unfamiliar to the subjects. All participants were tested individually. All of them were naïve and their visual ability was normal or corrected-to-normal as was the case in all subsequent studies and experiments.

Material. Photographs of 22 female students from the DADA-O-Faces database (Carbon, 2001) were used. Among them were no students of psychology at the Freie Universität Berlin, in order to reduce the probability that test subjects would be familiar with them. This procedure prevents any confounding of the task with earlier experience with this material. The mean age of the photographed persons was 23.4 years with a range from 20 to 34, which corresponds to the age of the participants in the study. They wore no eyeglasses or tattoos and the expression on their faces was neutral to further prevent any problems of expressions and the differences in recognizability of expressive faces.

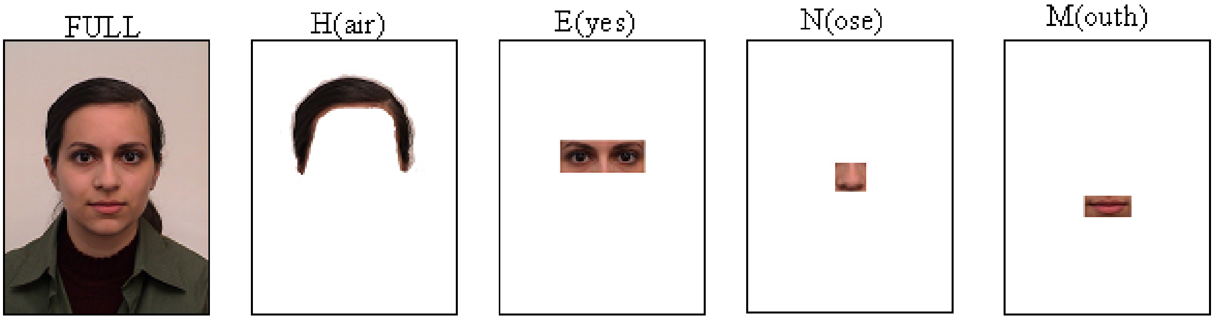
The photos were taken with a Mini-DV-still camera (Sony DCR PC-110) equipped with a Carl Zeiss Vario-Sonnar T* Lens at a resolution of 864 x 1152 pixels (see sketch of the room where the photos were shot in Figure 8-1 in the appendix). The models had to look straight to the middle of the optical lens with a neutral but inartificial sight to obtain frontal images. These pictures were reduced to a resolution of 768 x 576 pixels with 2^{16} colors. Examples of these pictures are shown in Figure 4-1.

Figure 4-1: Six faces from the set of 22 faces. Here, face F01-F06. The full set of faces is shown in Figure 8-2 in the appendix).



To investigate the distinctiveness of the whole and the facial parts (eyes, nose, mouth, hair) of the faces, full version and four partial versions (with only the eyes, nose, mouth, hair) of every face were produced.

Figure 4-2: Example of one face (face F12) in its full appearance (most left) and its featural presentations (hair, eyes, nose, and mouth).



Procedure. The experiment was conducted on a Macintosh G4-350 with an external 21″–Nokia 445Xi CRT-screen. The resolution of the monitor was 1280 x 1024 pixels with a 85 Hz-refresh rate (non-interlaced), resulting in an averaged size of the (full) faces of 18 x 22.5 cm or a visual angle of 16.7° x 18.6° with subjects sitting 70 cm away from the front of the screen. The luminance of the screen was 160 cd/m^2 (all luminance specifications were measured by a Gossen Mastersix photometer) (cf. Halsted, 1993). The experimental control was realized like all other following experiments through PsyScope PPC 1.25 (Cohen, MacWhinney, Flatt, & Provost, 1993), which allows the presentation of the stimuli within one single CRT-refresh cycle.

The study phase consisted of four separated blocks with short breaks between them. The first block was split into five sub-blocks: First, the participants had to rate the distinctiveness of the complete set of the whole faces on a *Likert scale* (Likert, 1932) from one to seven (from very indistinctive to very distinctive)[36] by pressing buttons on a keyboard. Afterwards, they saw all the single eyes, noses, mouths, and hairstyles in sub-blocks. All stimuli were randomized within each sub-block. The instruction for the rating (‘How easily would the face stand out in a crowd’) was adopted from previous studies about distinctiveness (Valentine & Bruce, 1986a, 1986c).

In the second block, the subjects’ task was to point with the mouse cursor on the most distinctive part of the faces. For each feature, all *Hotspots* that were located on this feature were added up to get a measure of distinctiveness for each of these features. This procedure (referred to as *Hotspot* hereafter) allows a very intuitive and easy method for showing distinctive areas in complex stimulus patterns without the need for verbalization.

In the third block, the participants had to rate the attractiveness of the whole faces on a scale from one to seven (from very unattractive to very attractive), as in several studies by Locher et al. (1993).

In the fourth and last block, subjects were given a similarity decision task. All possible face pairs of all the 22 faces were presented one after another (231 pairs total). As in block 1, a seven-point scale with the same valence orientation was used. This is a relatively small scale range compared with other studies, e.g. Sergent (1984b), who used 25 points, but it seemed to be more ergonomic for the participants. A speeded decision task was used, especially for the similarity-rating, because a speeded sequential judgments-task may provide better MDS spaces (Busey & Tunnicliff, 1999). The participants needed about 40-60 minutes for the whole procedure. They were tested individually. The order of stimuli within each block was randomized.

Results & Discussion.

Distinctiveness of features

The mean distinctiveness data, sampled for all subjects, of the different partial eyes, nose, mouth, and hair for every face are shown in Table 8-2 in the appendix. Due to a focus on inner facial features (definition of inner/outer features is illustrated in Figure 4-3) in the subsequent experiments, only these features (eyes, nose, mouth) will be considered.

---

[36] Some researchers prefer ratings on a unidimensional scale from ‘very typical’ to ‘very distinctive’ (e.g., Johnston et al., 1997), but a continuum from typical to distinctive is not unproblematic, because of its implicit assumption that typicality is the diametric opposite of distinctiveness (Vokey & Read, 1992). Therefore, I prefer the use of a continuum from ‘very indistinctive’ to ‘very distinctive’.

Figure 4-3: Illustration of the definition of inner and outer features used in the present work. Inner features are in particular the eyes, the nose, and the mouth. Outer features are the face contours, especially the hair region.

The distinctiveness ratings were split into two discrete groups for every feature. The *low distinctiveness* group contained the lower half of all distinctiveness ratings cut by the median of all scores, the *high distinctiveness* group comprised the upper half. Table 4-1 lists all features of the 22 original faces from Pre-Study 1a classified according to their distinctiveness. Within this table, bold names indicate features, which were used for Pre-Study 2.

Table 4-1: Facial parts of low and high distinctiveness. Bold names indicate features, which were later used for *DADA-M-faces* in Pre-Study 2.

	Low distinctiveness feature pool	High distinctiveness feature pool
Eyes	**F02**, **F05**, **F09**, **F01**, F17, F19, F03, F04, F06, F21, F14	F15, F08, F11, F18, F22, F20, F13, **F16**, **F07**, **F10**, **F12**
Nose	**F19**, **F09**, **F20**, **F10**, F08, F12, F06, F14, F05, F21, F15	F17, F01, F03, F07, F13, F04, F22, **F18**, **F02**, **F11**, **F16**
Mouth	**F21**, **F02**, **F17**, **F22**, F03, F06, F18, F07, F05, F13, F01	F19, F20, F08, F09, F15, F04, F14, **F12**, **F10**, **F11**, **F16**

The inter-rater reliability was calculated by Cronbach's α, which indicates how well a set of items (or variables) measures a single unidimensional latent construct. If data have a multidimensional structure, Cronbach's α will usually be low. Technically speaking, Cronbach's α is not a statistical test—it is a coefficient of reliability (or consistency) (SPSS Inc., 2001; Cortina, 1993). The consistency of the attractiveness was very high: .924, as was the consistency of the distinctiveness ratings for full faces (.787). Table 4-2 gives an overview of all Cronbach's α values, revealing that the consistency of eyes-ratings was relatively low.

Table 4-2: Consistency of the subjects' ratings (measured with Cronbach's α: N of cases: 22, N of items: 16); all non-relevant data for the following experiment appears in gray.

Task	*Cronbach's* α
Attractiveness (full faces)	.924
Distinctiveness (eyes)	.592
Distinctiveness (nose)	.680
Distinctiveness (mouth)	.818
Distinctiveness (hair)	.884
Distinctiveness (full faces)	.787

The distinctiveness consistency was rather high, but still lower than the one found by Valentine and Endo (1992) or Lee, Byatt, and Rhodes (2000), who found a Cronbach's α of .93.

To examine which of the inner parts (eyes, nose, and mouth; for mean distinctiveness ratings see Table 4-3) was the most distinct, a one-way repeated measurement ANOVA with the within-subjects factor of FEATURE (eyes, nose, and mouth) was calculated. This factor was significant ($F_{2,30}$=15.58, P<.0001; see Table 8-3 in the appendix), with all levels being significant against each other via Scheffé post-hoc tests (Ps<.0332; see Table 8-4 in the appendix).

Table 4-3: Mean Distinctiveness ratings for the inner parts and the full faces.

Task	*mean distinctiveness rating*
Distinctiveness (eyes)	3.702
Distinctiveness (nose)	2.710
Distinctiveness (mouth)	3.210
Distinctiveness (hair)	3.048
Distinctiveness (full faces)	3.463

Predictors for distinctiveness

Distinctiveness is a central variable for recognition processes. Therefore, it is important to know which attributes of a face predict the overall distinctiveness. In the following section, some possible predictors are discussed, mainly distinctiveness of facial parts and attractiveness.

The interrelation of the distinctiveness ratings for full faces and the ratings for the inner parts was tested in a multiple linear regression analysis. The analysis revealed a significant linear relation ($F_{3,21}$=4.063, P=.0228) with a relatively high correlation of R=.635. A further correlation analysis revealed that especially the eyes had a large impact on the ratings of the distinctiveness of full faces (R=.610; see Table 4-4 for more details).

Table 4-4: Triangle correlation matrix of the distinctiveness ratings.

	Full faces	Eyes	Nose	Mouth
Full faces	1.000			
Eyes	.610	1.000		
Nose	.167	.007	1.000	
Mouth	.425	.546	.226	1.000

The weakest predictor of the inner face parts for the distinctiveness of full faces was the distinctiveness ratings of noses: With a R of only .167, thus less then 3% ($.167^2$) of the distinctiveness variance can be predicted by the distinctiveness ratings of the noses.

Another possible predictor for overall distinctiveness might be attractiveness. According to the *averageness hypothesis* (Langlois & Roggman, 1990), average or medium distinctive faces are generally found to be highly attractive. Other researchers, on the opposite, assume that deviations from averageness lead to an increase in attractiveness in both male and females faces (e.g., Perrett, May, & Yoshikawa, 1994). Thus, although averageness is certainly attractive, it can be bettered by consistent deviations from averageness. For instance, Pollard, Shepherd, and Shepherd (1999) used facial-metric measurements of feature size and found that faces with features close to the mean size of the population studied were rated as average, rather than high, in attractiveness.

The consistency of inter-subjects ratings of attractiveness was, as is usually the case (Hönekopp, under review), very high (Cronbach's α=.924; see also Table 4-2). Moreover, the correlation between distinctiveness and attractiveness ratings, both for full faces, was only modest (*R*=.490), with a linear relation between the two measures (revealed by an ANOVA with $F_{1,20}$=6.321, *P*=.0206). The essential part of this relation depends on the eyes region (see Table 4-5), while the mouth and nose are scarcely involved.

Table 4-5: Correlation matrix of attractiveness rating for full faces and distinctiveness rating for part-based presented versions.

Feature	*Correlation with attractiveness*
Eyes	.580
Nose	.115
Mouth	.096

This is partly in accord with an experimental study in which only for male but not for female faces a strong relationship was found between face recognizability and attractiveness. Recognizability is itself strongly related with distinctiveness (O'Toole et al., 1998).

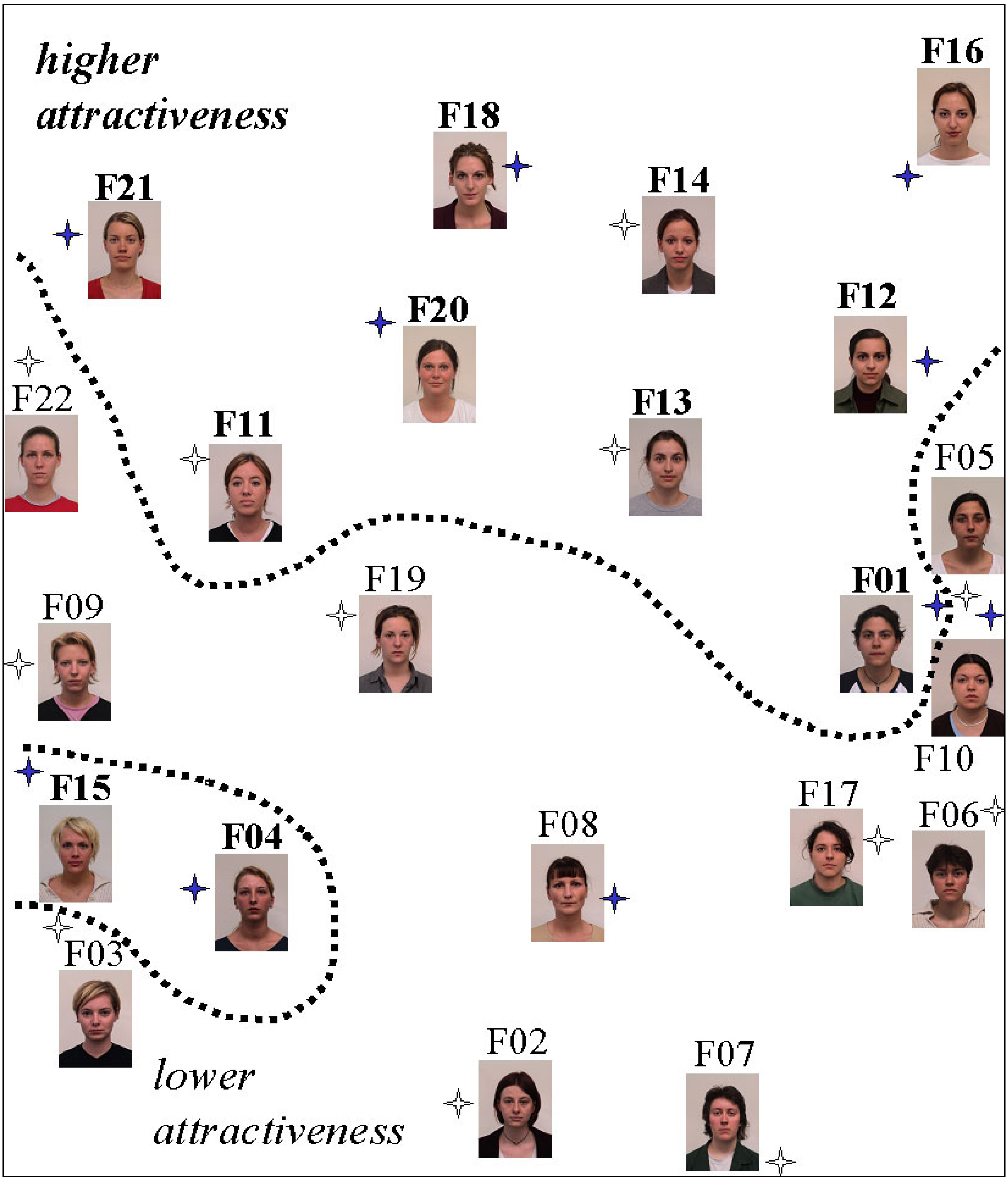
Face space

Another important predictor for distinctiveness might be the inter-similarity between faces. According to Valentine (1999), faces are represented in a *face space*. Within the psychological face space, more distinctive faces are situated in the periphery and typical faces rather in the centered area (Valentine, 1991; Lee et al., 2000; Johnston et al., 1997)[37]. To test this assumption, a two-dimensional metric MDS was calculated. The following results were obtained with a PROXSCAL MDS (see SPSS Inc., 2001) algorithm with an Euclidean metric. The MDS algorithm of the statistical package SYSTAT 5 (SYSTAT, 1992) with the same base parameters led to a very similar face space, thus validating the PROXSCAL solution.

---

[37] Lee et al. (2000) used 16 faces, evaluated by 24 subjects; Johnston et al. (1997) used 36 faces, evaluated by 12 viewers.

Figure 4-4: MDS solution for the similarity ratings (PROXSCAL with a 2D-euclidean space). Dark stars indicate highly distinctive faces; light stars, in contrast, indicate less distinctive ones. Additionally, highly attractive faces are symbolized by bold face names. The dotted lines cut the faces of higher attractiveness area from the area containing faces of lower attractiveness.



It is obvious from Figure 4-4 that faces of higher[38] attractiveness can easily be separated from faces of lower attractiveness, with two exceptions (Face F04 and F15). This in part contradicts the findings of Langlois and Roggmann (1990), who found attractive faces to be the mathematical average of faces in a population that are situated more in the center area of face space (Langlois, Roggmann, & Musselmann, 1994). Nevertheless, attractiveness had a great influ-

---

[38] Faces of higher or lower attractive, or of higher or lower distinctiveness, were identified by the median cut-off point of their related scale.

ence on the organization of the face space obtained by similarity ratings. This was also shown by many studies in the past. Therefore, attractiveness seems to be a very important social cue. Recently, it was even found that attractive children and adults are treated more positively than unattractive ones, even by those who know them (Langlois et al., 2000)!

In contrast, the metaphor of clear spatial separation of more and less distinctive faces is much more critical. Many highly distinctive faces—according to Valentine's (1999) model—are indeed located in the periphery of the face space, but Face F08 does not fit into this scheme, nor do some faces of lower distinctiveness in the lower parts of face space. Nevertheless, what seems apparent is the influence of hair color on similarity ratings. Blonde hair faces are clearly located in the lower left area, deep brown or black hair types in the right section and red or lighter brown females in the upper left region.

Hotspots

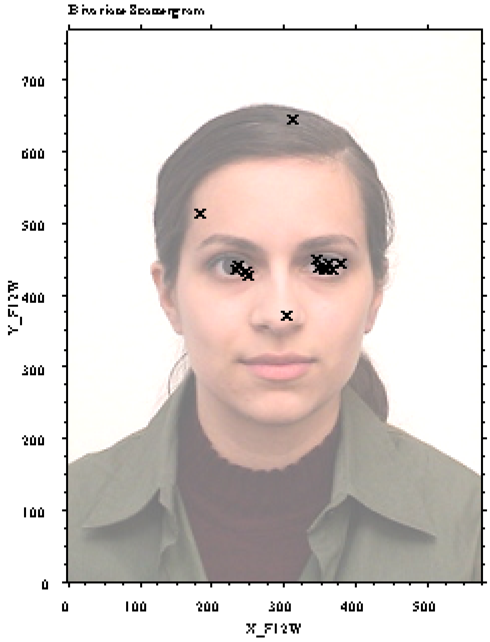
In this Pre-Study, a new distinctiveness measure called *Hotspot* was introduced. The logic of this measure is to mark the most distinctive area of a face. The participants had to click with the mouse cursor on the most distinctive point/area of a face. All hits on each feature were added up. Analogous to distinctiveness ratings, a similar distinctiveness pattern emerged for the inner features. The eyes were the most distinctive feature, the mouth the second most distinctive and the nose was the least distinctive area (see Table 4-6). This result was attained by adding up the *Hotspot* selections as described above. A one-way repeated measurement ANOVA with this counting of *Hotspots* as dependent measure and FEATURE (eyes, nose, mouth) as the within-subjects factor showed a significant factor FEATURE ($F_{2,30}$=13.52, P<.0001; see Table 8-5 in the appendix).

Table 4-6: *Hotspot* counting (percentage of all counting) for inner facial parts.

Task	*Hotspot counting (%)*
Eyes	30.7
Nose	9.9
Mouth	21.9
Within the faces (measure validation)	98.6

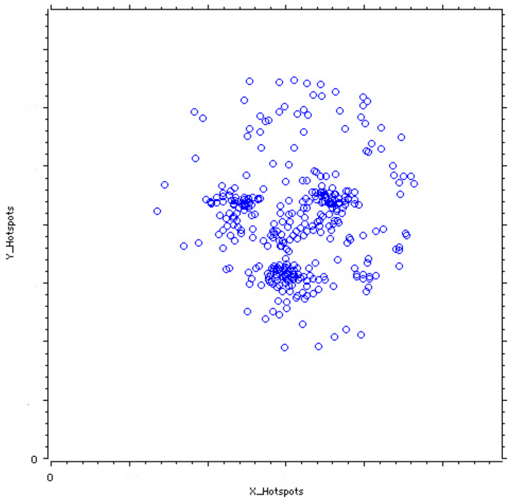
A Scheffé post hoc test (see Table 8-6 in the appendix), revealed a similar pattern to that of the distinctiveness data. Again, the nose was least distinctive, but now the difference between the eyes and the mouth was not significant in a very conservative Scheffé post hoc, while it was significant in a more liberal Fisher's PLSD post hoc test (P<.0359). 98.6% of all *Hotspots* lay within a face, which means that only 1.4% of the *Hotspots* were absurdly lying outside of a face. The amount of *Hotspots* lying outside of the face is an indicator for nonsense responses, because the task was to click on the most distinctive point *within* the face. With less than 2%, this indicator is negligibly low. Moreover, the *Hotspot* analysis for every single face further revealed that the inter-subjects reliability to click on the same spot of one face was rather impressive. A paradigmatical face (here face F12) with a high consistency of *Hotspots* will be found in Figure 4-5. Only one participant clicked on the nose and two on the hair region, but all other on one of the two eyes. The *Hotspot* data for all the other faces are shown in Figure 8-8 in the appendix.

Figure 4-5: *Hotspots* for face F12. Luminance and contrast has been adapted for a better illustration of the *Hotspots*.



Summarizing these findings, participants identified the eyes region in both procedures as the most distinctive part of the face. The nose was the least significant area and the mouth was of medium distinctiveness. A demonstration illustrating which areas are most preferred as highly distinctive parts is shown in Figure 4-6. In this figure, all *Hotspots* of every participant for all faces are shown in one single plot.

Figure 4-6: The ‚meta-face', which is constructed by placing all *Hotspot* points of every participant for every face together into one plot.



It is indeed easy to recognize a 'meta-face' at once, even though Figure 4-6 only consists of the most significant points of all 22 stimuli. The significant clustering of *Hotspots* in the eyes- and mouth-region in connection with sparse *Hotspot* data for the outline and the nose create an image resembling a face. Interestingly, the right side of this face is more accentuated (67.5%; two-tailed *T*-test against a base level of .5: *T*(15)=4.720, *P*=.0003), despite the fact that subjects were free to choose areas on the right or left side. This effect of the preference of the right side was not due to individual participants, but proved to be a consistent pattern of responses over nearly all subjects (only three participants selected *Hotspots* equally distributed over the left and right side, while all other participants preferred the right side). Was this a psychological effect of generally preferring the right side of a face when asked for distinctive points, or is it simply an artifact due to asymmetrical faces or light conditions? The first possibility seems to be rather counterintuitive, because as Gilbert and Bakan (1973) and Burt and Perrett (1997) have demonstrated, observers unknowingly tend to overestimate the importance of information from the left side of the face, but not from the right side as has happened in this Pre-Study. For an illustration of this effect see Bruce and Young (1998). In order to test the second possibility, namely that this effect is only an artifact due to the presented material, Pre-Study 1b was launched. In Pre-Study 1b, mirror-reversed versions of the stimuli used in this study were employed. See the next paragraph for more details.

In a further analysis, the *Hotspot* data were compared with the distinctiveness ratings. Both measures revealed the same trend of distinctiveness order within the inner parts of a face. How strong is the interrelation between these two measures? Do they both measure the same kind of distinctiveness or are they useful tools to investigate different psychological phenomena? In order to test this, the number of *Hotspots* was related to the distinctiveness rating. In this context, a positive correlation means that a high distinctiveness rating corresponds to a high amount of *Hotspots* selected for this feature. To test the relationship between these two measures for the inner features, three different correlation analyses were calculated. Table 4-7 shows the correlation indices for distinctiveness ratings and *Hotspot* data split by features.

Table 4-7: Correlation matrix of distinctiveness rating and *Hotspot* data split by features.

Feature	*Correlation*
Eyes	.368
Nose	.292
Mouth	.406

Table 4-7 shows that there is only a weak correlation between the two measures, but it seems that the *Hotspot* procedure does not only reflect the rating measure of distinctiveness, but is an autonomous measure. Nevertheless, it is rather problematic to compare both measures. Distinctiveness ratings are a graded measure of the *overall* distinctiveness of a face, whereas *Hotspot* data indicate the *position* of the most distinctive area of a face. Therefore, only further investigations could reveal the character and specific application of this new and simple measure, but this approach was not followed in the present study.

4.3.2 Pre-Study 1b: Mirrored Faces

Introduction.

Pre-Study 1b is an expansion of Pre-Study 1a that further investigates the right side preference in the *Hotspot*-task.

Side preference in Hotspot-task

In Pre-Study 1a, it was found that within the so-called *Hotspot*-routine, the participants showed a significant tendency of clicking on the right side of the shown faces more often than on the left side. The reason for this side preference will be scrutinized. Therefore, the *Hotspot* task from Pre-Study 1a was again used, but this time with mirror-reversed pictures. This enables us to check if this ‘effect’ is only a pictorial artifact of specific photographic pictures. If there is a general preference of the right side when people are asked to click on the most distinctive point of a face, then we are dealing with a psychologically relevant phenomenon. On the other hand, if it is only a material artifact, the left side has to be preferred with mirrored stimuli.

Method.

Participants. Sixteen students participated in the experiment (14 female; average age: 23.9 years). None of them had participated in Pre-Study 1a.

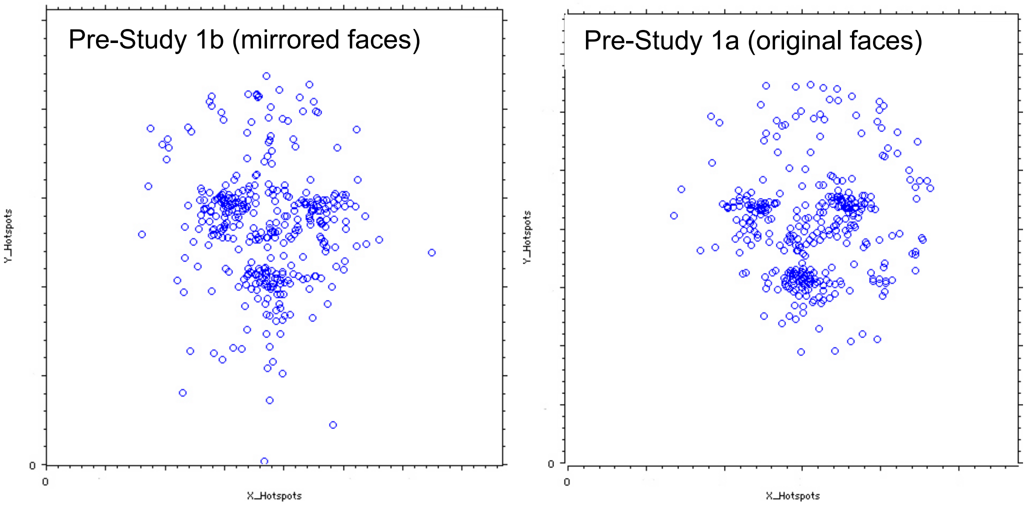
Material. The same photographs of 22 female students from Pre-Study 1a were used, but this time mirrored. Thus, the left side is now on the right side and vice versa.

Procedure. The procedure was the same as the *Hotspot*-task in Pre-Study 1a.

Results & Discussion.

To test the hypothesized general preference of right distinctive points, the number of clicks on the left and the right side was compared. In comparison to Pre-Study 1a, this time the subjects had a tendency to click on the left side of the faces (59.7% of all *Hotspots* lay on the left side; two-tailed *T*-test against a base level of .5: *T*(15)=2.15, *P*=.0487). Figure 4-7 gives an illustration of all *Hotspots* for all faces combined in form of a 'meta-face'.

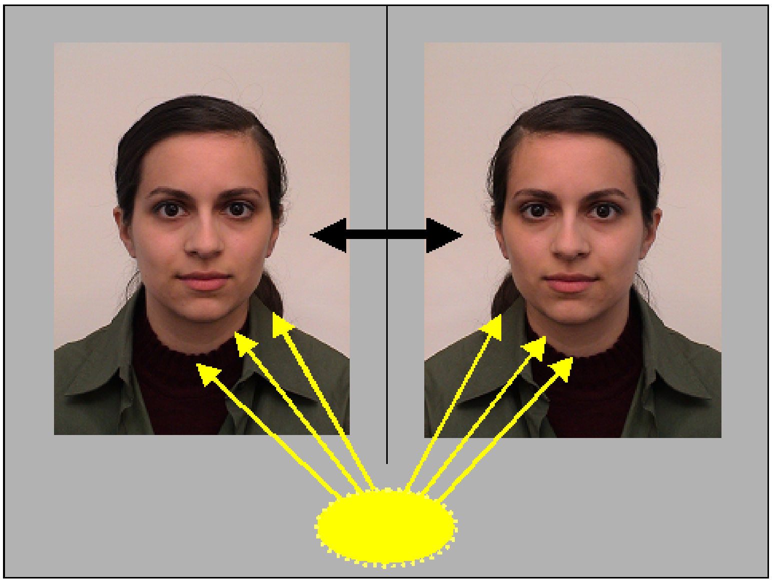
Figure 4-7: ‚Meta-faces'. On the left, the *meta-face* built from the *Hotspot* data of Pre-Study 1b with mirrored faces; on the right, the *meta-face* from Pre-Study 1a.



The 'meta-face' of Pre-Study 1b also seems to be the mirrored version of that given by the *Hotspots* of the original Pre-Study 1a. This tendency was once again not generated by individual participants but by nearly all of them. The pictorial asymmetry effect was further examined by analyzing the stimuli. A tiny asymmetry in the lighting of the photographed heads was revealed as a possible cause of this tendency. Consequently, an appraisal of the room in which the photographs had been taken (Figure 8-1 in the appendix) revealed a stronger light source shining onto the right side of the faces (and therefore onto the right side of the pictures in Pre-Study 1a, and onto the left side of the mirrored pictures in Pre-Study 1b).

This lighting effect is illustrated by Figure 4-8. Thus, from the given material it can be concluded that a more brightly lit part of a face seems to be more distinctive (in terms of *Hotspots*), perhaps due to greater contrasts.

Figure 4-8: Testing the mirrored stimulus material from Pre-Study 1a to exclude any psychologically relevant hypotheses about the right-clicking tendency in the *Hotspot* procedure (left: typical face from Pre-Study 1a; right: mirrored material from Pre-Study 1b). The origin of the main light source can easily be seen; it is symbolized by a yellow ellipse with rays reaching to the preferred areas.



In the subsequent experiments, a priori facial asymmetries are not very problematic. Nevertheless, the lighting has to be controlled for other studies, where the side of the face is theoretically essential.

4.3.3 Pre-Study 2: Artificially constructed Faces

Introduction.
There were three main goals of Pre-Study 2.

Construction of faces

First, artificial faces had to be constructed by combining systematically selected facial parts from Pre-Study 1a. This was done to systematically test the impact of face regions that were changed in different ways on the recognition process. The main kinds of manipulation were local (LOCAL) and configural (CONFIGURAL) alterations. See section *Material* for further details.

Evaluation of distinctiveness, attractiveness, and plausibility

Second, the participants had to evaluate this material in respect of its distinctiveness, attractiveness, and plausibility. Distinctiveness is a valid predictor for recognizability and therefore it is important to control the distinctiveness for testing recognition processing hypotheses. Additionally, attractiveness and plausibility ratings are analyzed to evaluate the quality and naturalness of the artificial faces. Ideally, the two newly created face classes CONFIGURAL and LOCAL should not differ from each other in terms of attractiveness and distinctiveness, in order to exclude any CLASS distinctiveness artifacts that might influence explanations of RT and correctness data in the succeeding main experiments (Exp.1 and Exp.2).

Face space

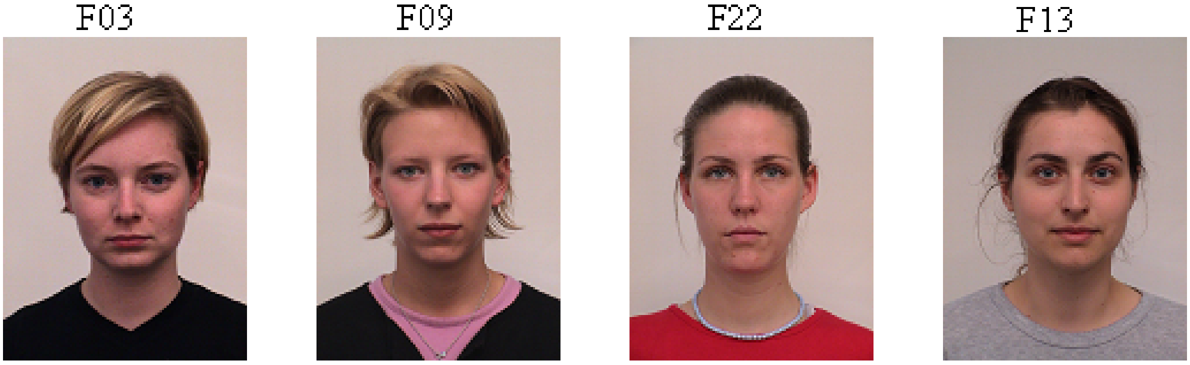
Third, it was analyzed how the psychological face space interconnects with the distinctiveness ratings. Therefore, a visualization of the face space had to be generated by calculating the similarity ratings of all possible face pairs. The psychological face space was measured by asking subjects to conduct pair-wise similarity ratings between all pairs of faces. The results have been submitted to a multidimensional scaling (MDS) procedure (Busey, 1998), just like the routine used in Pre-Study 1a.

Method.

Participants. Sixteen students participated in the experiment. The participants were undergraduate students (11 women, 5 men) of the Freie Universität Berlin, who were given a course credit to fulfill course requirements. The mean age was 26.1 years (21-36 years). All participants had normal vision abilities or were corrected to normal vision and none of them had participated in one of the former studies.

Material. The facial parts evaluated in Pre-Study 1a were separated into two feature pools according to their distinctiveness, a less distinctive (low distinctiveness) and a more distinctive (high distinctiveness) group. Four faces (face F03, F09, F22, and F13) were selected to represent the outline of the *basic faces* (see these faces in Figure 4-9).

Figure 4-9: Original faces (DADA-O-Faces) used as 'outlines' for the new faces to be constructed DADA-M-Face set.



F03

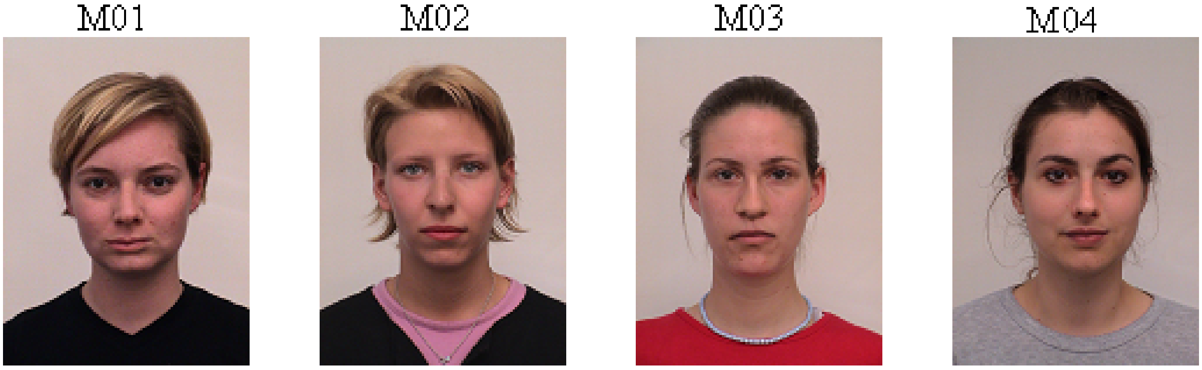
F09

F22

F13

Then four *basic-faces*, called M01, M02, M03, and M04 were created. Figure 4-10 shows these artificial faces. All cardinal inner features such as eyes, noses, and mouths have been replaced for M-faces. This can be seen by comparing the inner part of the faces in Figure 4-9 and Figure 4-10.

Figure 4-10: *Basic faces* of the DADA-M-Face set; 'M'-Faces means that these faces are manipulated ones.



M01

M02

M03

M04

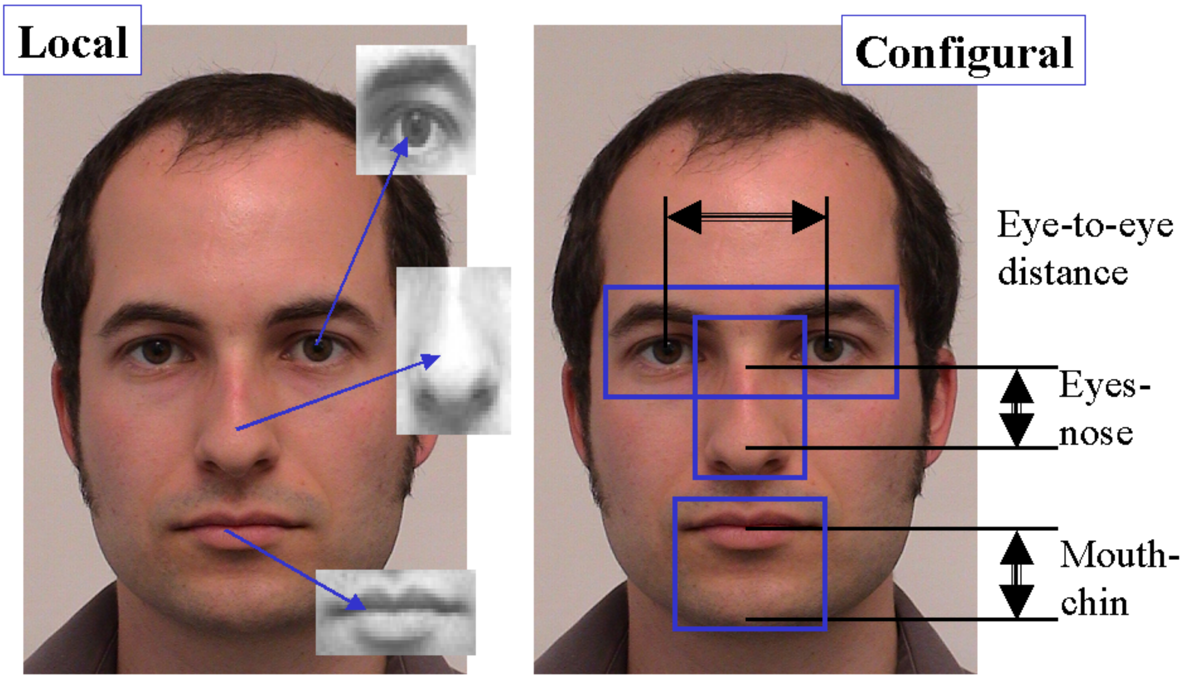
Based on these faces, additional faces were constructed. There were generally two classes of manipulations that were made: Firstly, locally altered versions (LOCAL), secondly, configurally changed versions (CONFIGURAL). Both manipulation classes were quite similar to the design of Leder and Bruce (2000) and Rhodes et al. (1993)[39]. However, within both manipulation classes, a low and a high salient version of each face

---

[39] Configural changes correspond to the *R2* manipulation technique described in Rhodes et al. (1993); local changes correspond to the *F2* manipulation.

were generated. For the LOCAL version[40], this was achieved by substituting the original parts of the face by either a low or high distinctiveness member of the different facial parts pools (see Figure 8-4 in the appendix). In the CONFIGURAL version (see Figure 8-5 in the appendix), the essential parts (nose and mouth) were shifted downwards either in a small step (seven pixels) or in a larger step (14 pixels). Analogous to the mouth and the nose, for the eyes region this was achieved by moving the eyes the same distance inward (similar to the manipulation in Leder & Bruce, 2000b). All these manipulations were made for all three main face areas plus the combinations of them resulting in seven different versions. These were: only for the eyes area (E), the nose area (N), the mouth area (M), the eyes and the nose area (EN), the nose and the mouth area (NM), the eyes and the mouth area (EM) and for the eyes, the nose and the mouth area together (ENM). For an illustration which areas were involved in the manipulations, see Figure 4-11. For each of the faces this resulted in 2 [CLASS] x 2 [SALIENCE] x 7 [FEATURE] = 28 face versions plus the *basic face*. The key for all manipulations creating these artificial faces from the original face parts is shown in Table 8-1 in the appendix.

Figure 4-11: Manipulated regions of a face through the E, N or M alterations for both face classes. On the left, the *local* manipulation class is shown, where discrete facial parts were replaced; on the right, *configural* alterations as described above are shown.



Procedure. The course of events was similar to that of Pre-Study 1a. Artificial faces were the stimuli presented to the subjects. Each participant saw only one face basis with its 28 variants. Because there were four different face bases, every fourth subject got the same stimulus material.
An additional task at the end of the study was introduced. In this task, the participants had to rate the subjective plausibility of the presented faces. The plausibility was investigated by asking the participants the question "how strongly does this face resemble a human face?" (Cooper & Wojan, 2000). Since all the faces used in this study were manipulated artificial faces that were constructed from original faces, this procedure was important to test whether those faces were plausible enough to be identified as 'normal' faces.
The whole procedure took about 50 minutes. All participants were tested individually.

---

[40] Only rarely have face researchers used pure versions of either sort of information. Most often, local features consist of swapped features from other faces, as in this study (Rhodes et al., 1993; Tanaka & Farah, 1993). Although this routine might best be called *componential* or *local* changing, every local change also alters the spatial relationship to the other features. Therefore, it is not a *pure* local change (Leder et al., 2001; Leder & Carbon, subm.-a). However, to reveal maximally realistic faces, this method was used nonetheless.

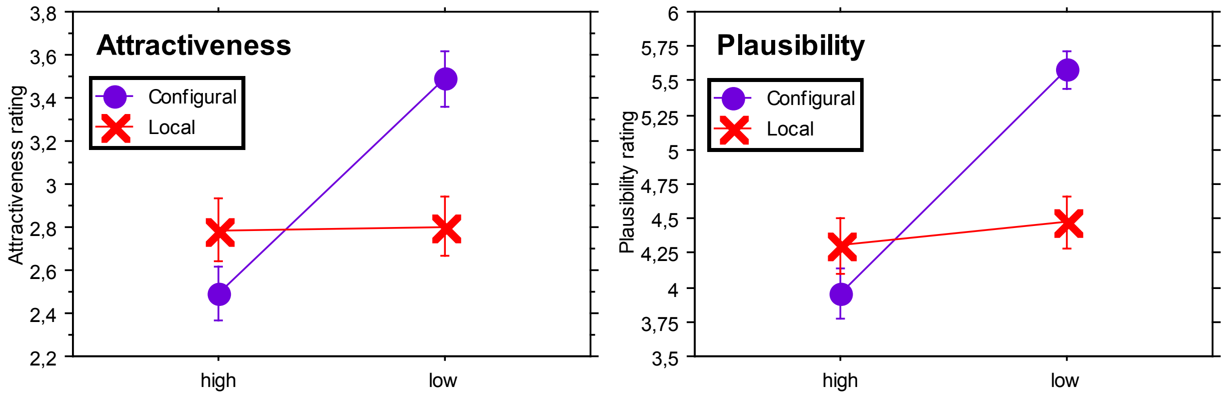
Results & Discussion.

Evaluation of distinctiveness, attractiveness, and plausibility

The raw data for ratings of distinctiveness, attractiveness, and plausibility are shown in the appendix (Table 8-7, Table 8-8, Table 8-9, respectively). First, it was tested whether the different manipulation classes (CLASS: configural vs. local) differ from each other in terms of attractiveness, distinctiveness, and plausibility. Any differences would be problematic for testing recognition data and processing hypotheses, because different degrees of distinctiveness, plausibility, and attractiveness are possible candidates for biased cognitive processes. Thus, it is important to exclude that material artifacts confound the subsequent experimental data. To test for possible differences, three different three-way repeated measurement ANOVAs (dependent measures: *attractiveness*, *distinctiveness*, and *plausibility*) with the following within-subjects factors were run: CLASS (configural vs. local), SALIENCE (low or high salient) and FEATURE (E:Eyes, N:Nose, M:Mouth, EN: Eyes&Nose, EM, ENM, NM). The detailed results of these ANOVAs will be found in the appendix (Table 8-10, Table 8-11, and Table 8-12, respectively). CLASS was not significant in any analysis ($F\text{s}_{1,15}<1.85$, $P\text{s}>.1945$). Thus, with the given statistical power, there was *no* difference between locally and configurally manipulated items in respect to these three measures. Normally, experiments in which local and configural faces are used as material neglect this test. Yet without such a test, it is unsure which factors are responsible for any differences in subsequent behavioral data. Here, both stimuli classes, configural and local, do *not* differ prior to the experiment, thus any discrepant behavioral data will be caused by specific processing and are not due to material artifacts.

Concerning the salience of stimuli, all analyses indicate significant main effects of SALIENCE ($F\text{s}_{1,15}>14.89$, $P\text{s}<.0016$). Highly salient faces were found to be more distinctive than faces of low salience. This validates the distinctiveness results of Pre-Study 1a. Moreover, high salience significantly reduced the plausibility of these faces. The influence of salience on attractiveness was very similar. Highly salient faces were evaluated as being less attractive. The combination of highly salient features might be uncommon for the beholder and thus reduces plausibility and attractiveness. Interestingly, this impact of salience on plausibility and attractiveness was only found for configural faces. As interactions between CLASS and SALIENCE revealed ($F\text{s}_{1,15}>14.59$, $P\text{s}<.0017$), only the salience of configural faces influenced plausibility and attractiveness. Figure 4-12 illustrates these findings.

Figure 4-12: Interaction between CLASS and SALIENCE on the variables *attractiveness* and *plausibility*. The x-axis refers to high and low salience. Error bars indicate the standard error (*SE*) of the mean as in all subsequent diagrams below.



If faces were altered configurally, the more the features were shifted, the less attractive and plausible these faces became. Natural faces also differ in terms of featural as well as configural aspects. Nevertheless, what seems very uncommon is a simultaneous alteration of many configural aspects. Since Galton's (1879) famous work on composite faces, it has become common sense that all faces differ only subtly in respect to their configuration. Average attractiveness correlates highly with averageness of configuration (Langlois & Roggman, 1990; Burt & Perrett, 1997). Highly distinctive configural faces are *not* average, because they deviate significantly from a 'central prototype' with its harmonious and overall average relations between the features. Therefore, such faces seem to be extreme entities, which are very uncommon for a natural face. This unnatural outlook reduces plausibility as well as attractiveness.

As in Pre-Study 1a, the nose was again rated as being least distinctive, but the distinctiveness of the eyes was not *generally* the highest rated feature (see Table 4-8).

Table 4-8: Distinctiveness data for the cardinal features E, N, and M, split by CLASS. The *SD*s are given in square brackets.

Distinctiveness	Configural	*SD*	Local	*SD*
E	3.219	[1.211]	3.719	[1.703]
N	2.500	[.753]	2.844	[1.351]
M	3.500	[1.238]	3.063	[1.094]
average	*3.073*	*[1.148]*	*3.208*	*[1.425]*

In order to test whether these differences were significant, two one-way repeated measurement ANOVAs were conducted, one for local and the other for configural faces. The only within factor was SINGLEFEATURE (E,N,M); the dependent variable was distinctiveness (see Table 8-13 and Table 8-14 in the appendix, respectively). Both analyses revealed a significant main effect SINGLEFEATURE ($Fs_{2,30}$>3.98, *P*s<.0294). Interestingly, Scheffé post-hoc tests for both ANOVAs revealed that only the difference between M and N for configural faces, and the difference between E and N for local faces were significant (see Table 8-15 and Table 8-16 in the appendix, respectively). Therefore, the eyes were not always the most distinctive feature as in Pre-Study 1a.

Furthermore, it seems that the three measures *distinctiveness*, *plausibility,* and *attractiveness* are highly inter-related. To quantify the inter-relationship, an analysis of intercorrelations was conducted. Therefore, the ratings averaged over all subjects of the three measures were used as dependent variables and the 28 manipulated faces as different cases. Plausibility and attractiveness seem to be nearly perfectly correlated. Additionally, both measures correlate highly negatively with distinctiveness. This means that a face consisting of highly distinctive features or configurations is unattractive and not very plausible for the subjects. Table 4-9 shows the intercorrelations of the three measures.

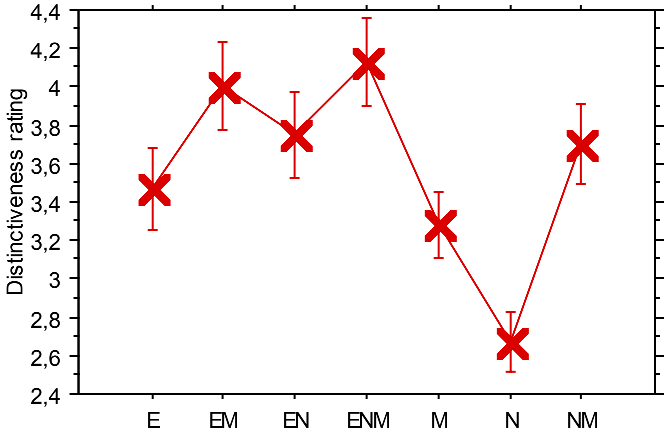
Table 4-9: Intercorrelations between the measures *attractiveness*, *plausibility*, and *distinctiveness*

Intercorrelations	*Attractiveness*	*Plausibility*	*Distinctiveness*
Attractiveness	1.000		
Plausibility	.868	1.000	
Distinctiveness	-.825	-.883	1.000

Most importantly for the following recognition experiments is the *distinctiveness* scale. Distinctiveness of features is an outstanding predictor for face recognition and the speed of its processing (Shapiro & Penrod, 1986). For the three-way ANOVA of distinctiveness discussed above, there were some more significant effects found, besides the SALIENCE ($F_{1,15}$=51.124, P<.0001) factor described above.

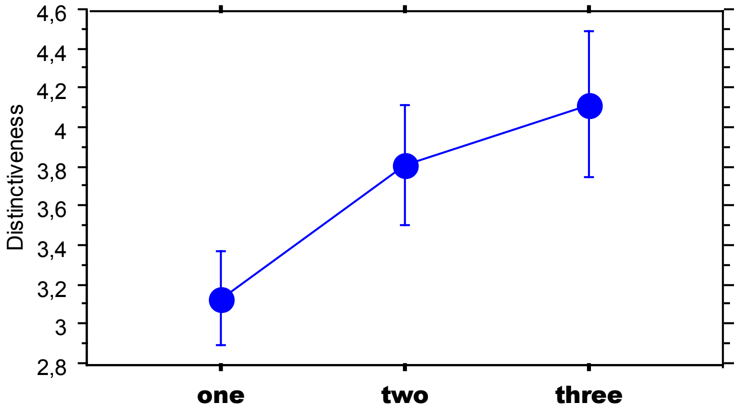
Significant were the main effect of FEATURE ($F_{6,90}$=10.281, P<.0001), the two-way interaction between SALIENCE and FEATURE ($F_{6,90}$=2.758, P=.0166) and the three-way interaction of CLASS, SALIENCE and FEATURE ($F_{6,90}$=2.455, P=.0303). Concerning the factor FEATURE, there is a clear dissimilarity between the different features (see Figure 4-13). It seems, that the more the items were changed, the higher their distinctiveness became.

Figure 4-13: Distinctiveness data ordered by FEATURE.



This hypothesis was tested with an additional repeated measurement ANOVA with NUMBER (1 [E,N,M], 2 [EN, EM, NM] or 3 [ENM] manipulated areas) as the only within factor and distinctiveness again as the dependent variable. The factor NUMBER emerged as significant ($F_{2,30}$=22.268, P<.0001; see Table 8-13 in the appendix), as is illustrated in Figure 4-14. Scheffé post-hoc tests revealed that only the difference between two and three areas changed simultaneously failed to reach significance (see Table 8-18 in the appendix).

Figure 4-14: Differences of distinctiveness depending upon the number of changed face areas (factor NUMBER).

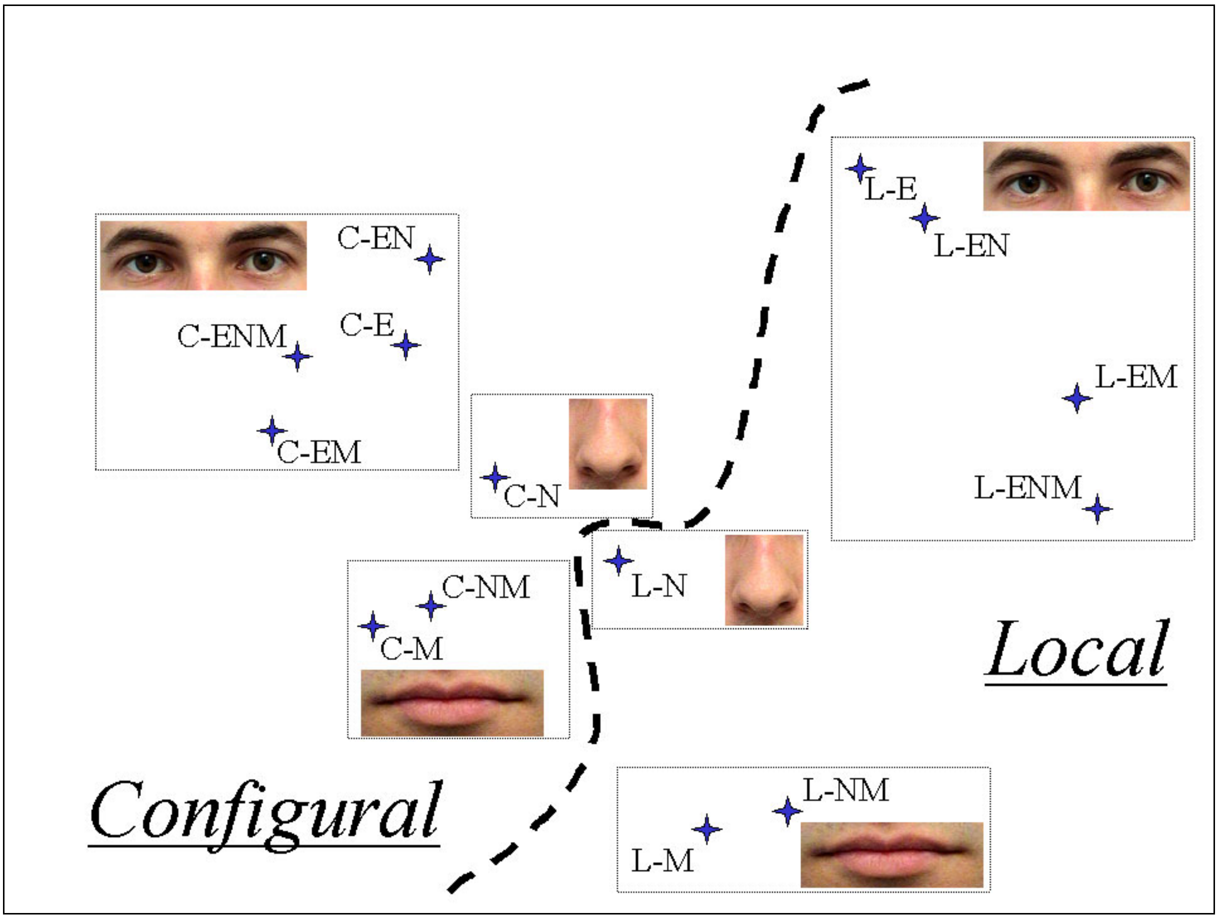


Within the single featural changes (E, N, and M), alterations to the nose were the least distinctive ones. This is in accord with perceptual data indicating that changes of the nose are sometimes hardly recognizable in frontally viewed faces (Haig, 1984; Matthews, 1978). One plausible reason for this might be the relatively low perceptual information content or the low contrast between a nose and the surrounding skin. Therefore, the specific figure-ground relation is hardly recognizable for frontal views[41]. It also seems reasonable that the relatively scarce social information of the nose might be a further factor (see Parks et al., 1985).

Face space

The second aim of this Pre-Study was to examine the face space spanned by the manipulated faces. The graphical animation of the face space is an illustrative tool to further analyze the similarity structure in a face set. As discussed in the introductory part of this study, similarity depends strongly on distinctiveness and is therefore an important predictor for face recognition. To calculate the face space, a PROXSCAL MDS (see SPSS Inc., 2001) with two dimensions and an Euclidean metric was used. For a similar modeling problem (Johnston et al., 1997), Euclidean metric has been applied successfully. The *Young S-Stress* was .06098 with an explained dispersion (D.A.F.) of .97541, which indicates a satisfactory transformation (Borg & Staufenbiel, 1997; Borg, 2000). Figure 4-15 illustrates the psychological face space generated by the systematically constructed DADA-M-Faces.

Figure 4-15: MDS solution for the complete similarity dataset of Pre-Study 2. Obviously, the locally and configurally manipulated faces are distinguished in two regions. Moreover, both classes were partitioned for an eyes facet, a mouth facet, and a nose facet (symbolized by an eyes, mouth, and nose icon, respectively).



This MDS solution seems not to be degenerated. This was revealed by the Shepard diagram in Figure 8-9 (in the appendix). The relation between transformed distances and the given distances has a relatively linear appearance without any obvious steps. Additionally, an alterna-

---

[41] This reflection is not valid for ¾- or profile views. Here, the nose is a very important feature due to its outstanding 3D curvature.

tive ALSCAL routine, which was proposed by other authors (Takane, Young, & De Leeuw, 1976; Young & Lewicky, 1980), revealed a very similar face space.

Figure 4-15 reveals that there is a clear separation of the local and configural CLASS (in this case, the configural on the left and the local on the right side, symbolized by a dotted line). Moreover, all changes in which the eyes area were involved are within a dense region. The same holds true for any mouth changes—as long as no eyes were involved. Contrary to this, the spatial positioning of the nose seems to be in accord with the low distinctiveness rating of noses found above and in Pre-Study 1a. An additional change of nose made no great difference for similarity data (i.e., EN vs. E, ENM vs. EM, NM vs. M). With the exception of local nose changes, the nose seems to play no important role in similarity evaluation.

This MDS solution is also in accord with Valentine's (1991) assumption of a face space of polar organization with highly distinctive faces lying in the outer zone and more prototypical in the central area (see an illustration of this assumption in Figure 3-1 of section 3.1). In this case, faces that are only changed in the nose area are being located in the central region. This is important to note, because faces found in the central region are probably harder to discriminate and therefore harder to recognize. Moreover, it is most likely that even their processing lasts longer or starts later, because of the high difficulty of discrimination.

The selection of only two dimensions at the same time seems to be mathematically unfavorable, because the *Scree*-Plot indicates that a four dimensional solution would be preferable (see Figure 8-10 in the appendix). However, a two-dimensional plot is more amenable to interpretation and comparison with other studies. It is plausible, as Valentine and colleagues have proposed, that the dimensions of the face space are acquired through a process of perceptual learning and are to be scaled to give optimal recognition performance for the population of faces given the current task requirement (Valentine, 1991; Valentine et al., 1995).

All these inferences could also be drawn from a subset of faces, which only consists of highly distinctive features. With an analogous PROXSCAL MDS once again there are clear facets identifiable with the nose as a very weak predictor for changes in similarity data. Figure 8-11 in the appendix demonstrates that even with highly distinctive noses, the appearance of the local and configural nose version is nearly equal. The distribution in the Shepard diagram and a *S-Stress* of only .03637 lends further support to the contention that the MDS solution was not degenerated.

Summary of the Pre-Studies

In Pre-Study 1a, natural faces were evaluated according to the distinctiveness of their facial parts. Additionally, several predictors for distinctiveness were tested. Attractiveness and the distinctiveness of the eyes were found to be important variables for the overall distinctiveness of a face. Eyes were found to be the most salient inner face feature, followed by the mouth. The nose, however, played no special role in predicting the distinctiveness of a full face. A newly introduced measure of distinctiveness called *Hotspot* showed high reliability and appears to be an autonomous measure besides the commonly used ratings.

The *Hotspot*-clicking tendency was biased by material artifacts. Participants clicked on more brightly illuminated areas with greater frequency. This artifact was exposed by Pre-Study 1b.

For Pre-Study 2, artificial faces were constructed systematically by combining the evaluated facial parts and facial outlines from Pre-Study 1a. Two major manipulations were conducted. First, faces were changed configurally by changing the inter-relation of the facial parts. Second, single features were replaced by alternative ones. The material was evaluated in respect to distinctiveness, attractiveness, and plausibility. The more the faces had been changed in a configural way, the less plausible and the less attractive these faces were rated. Moreover, it was again demonstrated that noses are not very salient and did not contribute much to the degree of idiosyncrasy of a face. This was revealed by distinctiveness data as well as by similarity ratings.

The artificial material constructed and controlled in these Pre-Studies will be used further for a highly systematized face recognition paradigm to investigate early processing of face recognition in Experiment 1 and Experiment 2.

4.3.4 Experiment 1: Processing of unfamiliar faces

Introduction.
Nearly 100 years ago, William James pondered how many ideas or things we can attend to at one time. His answer was "not easily more than one unless the processes are very habitual; but then two, or even three, without very much oscillation of the attention" (James, 1905, p.409).

Experiment 1 attempts to test this assumption in respect to face recognition with specific processing models. Testing of specific processing models needs specifically manipulated material. Therefore, highly standardized face material was used that was constructed out of the elements of the DADA faces (Carbon, 2001), evaluated in Pre-Study 1a (see paragraph 4.3.1). Configural as well as local manipulations (CLASS) were used (see for a description Pre-Study 2 in paragraph 4.3.3). Furthermore, different facial areas (eyes, nose, mouth) were manipulated. The combination of different presentation times (PTs) along with these systematic facial manipulations makes it possible to investigate the processing of different face areas. The use of a *feature substitution* paradigm in combination with photographic material in contrast to schematic faces (e.g. Fraser et al., 1990) improves the ecological validity of these experiments. Moreover, the special design of Experiment 1 allows for the testing of specific process-assumptions. The material was systematically manipulated in regard to local and configural features, which were controlled to have the same degree of distinctiveness. Additionally, the PTs were systematically varied to test the recognition performance under different time constraints.

Salience

In this experiment, it was investigated for which areas a perceiver is more or less sensitive; or, how distinctive these areas are for the cognitive system in terms of efficient early processing. DADA-M-Faces as described in Pre-Study 2 were used as stimulus material. Both saliency-versions, low and high salience, were used. The question is whether highly salient faces are processed more efficiently. The salience of a feature is assumed to be extremely important for the detection of a change. This should result in a main effect of salience and it should be discernible in recognition rates and in the speed of reaction, measured by RT (Massaro, 1989).

Feature dominance

Not only the overall distinctiveness of a face is assumed to influence the recognition processing. Moreover, there should be feature dominance of some areas. Alterations of some face areas are supposed to be more easily detectable than other areas. In respect to the inner facial elements, the eyes are the most important spot (Brunelli & Poggio, 1993). The mouth and the nose are probably less important than the eyes because they reveal less physical information. Nevertheless, in a frontal view the mouth is still more important than the nose, because a nose is most significant in a profile view, but not in frontal view. A mouth in the frontal view provides more physical and social cues than a nose (Ellis, 1975; Scassellati, 1998).

Feature quantity

Nevertheless, not only *single* features are believed to influence face recognition, but also the *combination* of these. It is plausible that the more features have been changed, the more likely it becomes that the alteration of the face will be detected (Smith & Nielsen, 1970). This would also be in accord with the distinctiveness results of Pre-Study 2. Moreover, a holistical account (e.g., Tanaka & Farah, 1993) would also assume that the more information is changed, the more the holistic *Gestalt* is altered. Therefore, the detection of several features should be more efficient when more features have been changed.

Validation of PT

In this experiment, the PTs were systematically varied. To test whether the setting of these PTs is effective, the recognition rate for each PT has to be analyzed. It is assumed that there is a general recognition advantage for longer PTs. If the PT range is well selected, there must be a significant effect of PT of the kind that the longer the original exposure to the target has been, the greater the probability of recognition becomes. If not, this is a hint for a PT range for which the cognitive apparatus is not sensitive with respect to face recognition. This is a direct validation of experimental design settings.

Rapid identification

The PTs used in this experiment were very short. With PTs from only 32 ms to 94 ms, there is not much time to process a complex stimulus like a face. Nevertheless, some researchers have found that it is possible to identify faces within about 100 ms of presentation (Biederman, 1981; Delorme, Richard et al., 1999; Locher et al., 1993). However, the use of varied PTs can reveal how much and which parts are recognized accurately.

Configural and local feature processing

In early face processing, different processing stages occur. The different main processing strategies are to detect configural and local characteristics, but both processes might not begin at the same time. However, configural processing might start later than local processing, because it is easier and faster to focus a single attentive feature than a greater configuration (Biederman, 1987)[42].

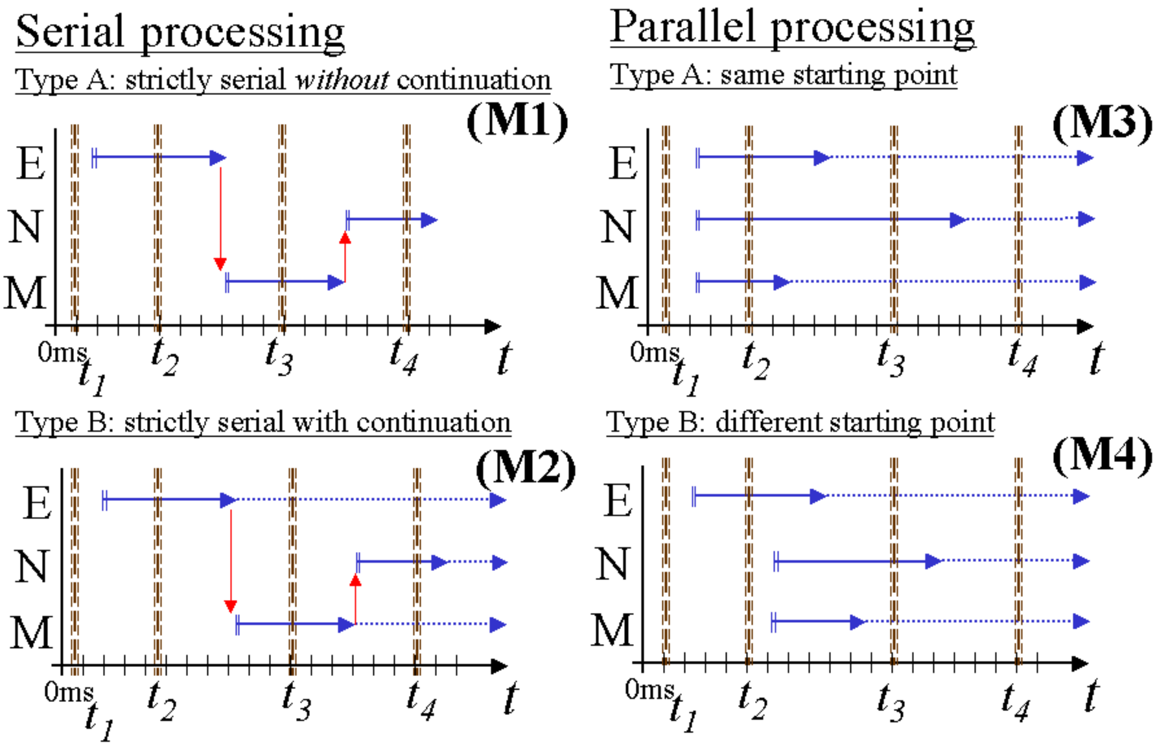
Feature processing

Additionally, this experiment analyzes the quality of the ongoing processes. Namely, whether they work in a serial or in a parallel manner. Furthermore, it is assumed that within each class, there is also a processing order of single features or regions. Analogous to the impact of salience discussed above, there is also an order in which different spots are triggered and therefore detected (Haig, 1984). According to this logic, the important locations should be recognizable at a prior state of processing than irrelevant or less informative ones, because it is essential to be able to identify most informative parts at first (Haig, 1986). An order of feature or location processing would support the microgenetic account (Flavell & Draguns, 1957). A measure that will be discussed later, called *WOM*, will investigate four different processing models. The first two of them represent serial accounts, the third and fourth model the class of parallel models. To test these models, the recognition rates for different PTs are calculated and compared with different processing models and their predictions. These models are illustrated in Figure 4-16. The detailed descriptions of these models are presented on page 94.

---

[42] Biederman's (1987) recognition-by-components (RBC) theory proposes a hierarchical set of processing stages, leading to object recognition. The identification of components of geons, which are local features, is thereby the first complex cognitive step.

Figure 4-16: Four different processing models. M1 and M2 with serial processing assumptions, M3 and M4 with parallel ones. For every model, the processing of the three cardinal inner features E (eyes) , N (nose), and M (mouse) is illustrated by blue arrows. t_1, t_2, t_3, and t_4 symbolize several PTs. The full description of these models is given on page 94.



Hypotheses.

Seven different hypotheses investigated the early face recognition and the ongoing processing. In order to test these hypotheses, three different measures were used. For testing accuracy, the correctness rates were analyzed. For testing the speed of processes, reaction times (RTs) and a new measure, called *WOM*, were investigated.

- Hyp.1–*Salience*: The salience of a feature is extremely important for the detection of a change. This should be evident in a main effect of salience as well as in recognition rates and in the speed of reaction (Massaro, 1989).

- Hyp.2–*Feature dominance*: There is feature dominance in that alterations in some face areas are more easily detectable than in others. Changes to eyes should be recognized best. Alterations to the mouth and the nose region should be detected less easily, with noses that are not easily detectable within the given PT-range (Ellis, 1975; Scassellati, 1998).

- Hyp.3–*Feature quantity*: The more features have changed, the more likely is the realization that the face was altered (Smith & Nielsen, 1970).

- Hyp.4–*Validation of PT*: General advantage for recognition rates for longer PTs.

- Hyp.5–*Rapid identification*: It is possible to process and identify faces even under a very short PT, here in under 100 ms (Biederman, 1981; Delorme, Richard et al., 1999).

- Hyp.6–*Configural and local feature processing*: Configural processing starts later than the processing of local features (Biederman, 1987).

- Hyp.7–*Feature processing*: Within each class, there is a processing order of single features or regions according to the microgenetic account (Flavell & Draguns, 1957). The more important locations should be recognizable at a prior state of processing than irrelevant or uninformative ones (Haig, 1986). Several specific processing assumptions, illustrated in Figure 4-16, are to be tested, most prominently serial and parallel processing types.

Method.

Participants. 28 students participated in the experiment. The participants were undergraduate students (18 women, 10 men) of the Freie Universität Berlin, who were given a course credit to fulfill course requirements. The mean age was 24.4 years (19-36 years). All participants had normal vision abilities or were corrected to normal vision and none of them had participated in one of the former studies to prevent any pre-knowledge of the presented face material.

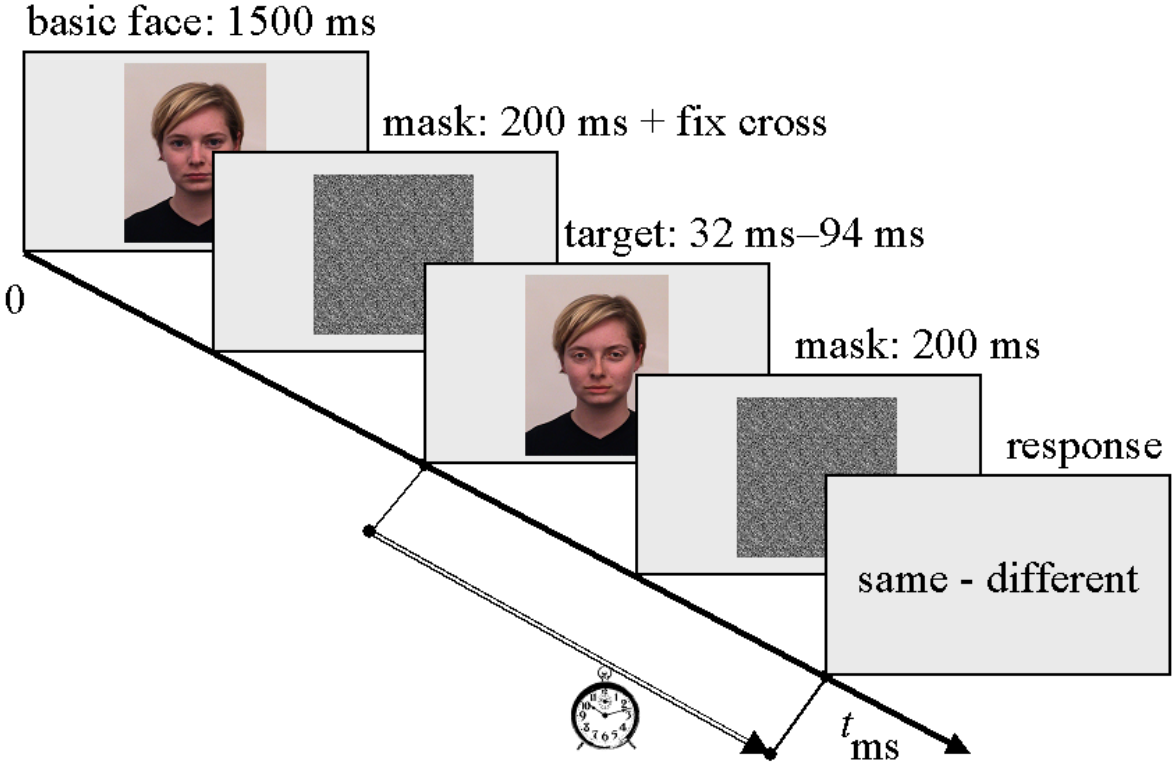
Material. The facial parts from the Pre-Study 1a (E:eyes, M:mouth, and N:nose) were split up in two same sized groups due to their averaged distinctiveness rating rank. The first group consisted of the low salient parts, the second of highly salient features. See for details in section 4.3.3.
The DADA-M-Faces, which are systematically constructed artificial faces, were used as stimuli. The construction of these faces is described in Pre-Study 2. The critical face feature areas that were changed were the nose, the mouth, and the eyes. To analyze all possible combinations of the locations, seven versions (factor FEATURE) were generated, consisting of all single features and the combinations of them (only eyes (E), only nose (N), only mouth (M), eyes and nose (EN), eyes and mouth (EM), nose and mouth (NM) and eyes, nose and mouth combined (ENM)). These location manipulations were combined with the factor SALIENCE (high or low salience manipulations), whereas alterations of ‘highly salient faces’ consisted only of highly-salient changes, evaluated as such by Pre-Study 1a. This was also the case for low-salience faces, consisting of low-salience changes. Additionally, for every face there was a configural or local version, indicating the manipulation class (CLASS). All these combinations result in 2 [CLASS] x 2 [SALIENCE] x 7 [FEATURE] = 28 possible alterations of every basic face, as described for Pre-Study 2.

Procedure. All subjects were tested individually. The experiment was conducted on a Macintosh Imac-350 (Mac-Os 9.0.4) with an integrated 15"-CRT-screen, situated 60 cm in front of the participants. The resolution of the monitor was 800 x 600 pixels with a 95 Hz-refresh rate with a size of the pictures of 288 x 384 pixels or a visual angle of 9.5° x 13°. The luminance of the screen was 220 cd/m^2. The experimental control was realized through the Mac-OS software PsyScope PPC 1.25 (Cohen et al., 1993), which allows the presentation of the stimuli within one CRT-refresh cycle. A *CMU ButtonBox* was used for the registration of the subject’s responses, allowing a measurement accuracy of one millisecond.
The participants were shown two facial pictures in sequence (first: basic-face; second: target), separated by a random mask composed of randomly distributed dots on a white field (as in Leder et al., 2001) with a duration of 200 ms (see illustration of this mask in Figure 8-6 in the appendix). In a *same-different task*, participants were asked to decide[43], whether the target and the basic-face were the same faces (Posner, 1969; McKenzie, Wixted, Noelle, & Gyurjyan, 2001). The response should be given in a fast but accurate way. The duration of the presentation of the basic-face was held constant through all conditions at 1500 ms, but the PT of the target varied in several steps (32 ms, 42 ms, 53 ms, 63 ms, 74 ms, 84 ms, 94 ms). After the presentation of the target, a second mask was also presented for 200 ms to prevent any afterimages. The PTs were set due to the refresh rate of the CRT-monitor (possible PTs were multiples of 1000/95 ms=10.53 ms). See Figure 4-17 for an illustration of the time-course of a trial.

---

[43] The assignment of the answers to the keys was balanced over the subjects as suggested by Neath (1998).

Figure 4-17: Time-Course of a trial in Experiment 1.



To balance all factors, a large number of trials were needed. Half the trials consisted of 'same', the other half of 'different' pairs. This resulted in the number of 2 [ANSWER: same/different] x 7 [PT] x 28 [manipulations] = 392 trials. Because of the large number of trials and the demanding nature of the task, there were three breaks during the test phase[44].

Further, in order to train the participants on this task, the participants were familiarized in a special training. The training phase used a particular training face which was manipulated in the same way as the test stimuli (on the basis of face F02 from Pre-Study 1a). For 14 training trials, the subjects received an acoustic feedback about the correctness of their answer; the PT for the targets in these training trials was fixed to 94 ms. The training trials and the first two trials of each test block were used as blindtrials and were excluded from data analyses.

Finally, the experimenter briefly interviewed the participants about the experiment. The participants had to report experiences and perceptions which they had made during the experiment. The total duration of this experiment including all the breaks and the post-experimental interview was about 50-60 minutes.

Results & Discussion.

The mean percentage correct and RT data of all correct trials were sampled over all subjects. Table 4-10 shows the full data for *accuracy* (percentage correct); Table 4-11 shows the *RT* data for all correct trials, respectively.

---

[44] This large number of trials is not uncommon for face recognition experiments. For instance, Bachmann (1991) used more than 1500 trials for each subject.

Table 4-10: Mean percent correctness for all experimental conditions. The PT conditions range from 32 ms (PT32) to 94 ms (PT94).

CLASS	SALIENCE	FEATURE	PT32	PT42	PT53	PT63	PT74	PT84	PT94
configural	high	E	0.71	0.71	0.79	0.64	0.86	0.86	0.71
		M	0.50	0.57	0.61	0.75	0.75	0.79	0.71
		N	0.54	0.61	0.54	0.71	0.61	0.61	0.57
		EM	0.82	0.89	0.89	0.96	0.96	0.89	0.96
		EN	0.75	0.75	0.82	0.93	0.89	0.93	0.96
		NM	0.68	0.75	0.79	0.71	0.68	0.79	0.93
		ENM	0.79	0.89	0.93	0.93	0.96	0.89	1.00
	low	E	0.36	0.43	0.61	0.79	0.79	0.79	0.57
		M	0.46	0.46	0.57	0.57	0.54	0.50	0.61
		N	0.25	0.36	0.43	0.54	0.54	0.50	0.39
		EM	0.50	0.79	0.64	0.82	0.82	0.64	0.79
		EN	0.54	0.68	0.71	0.61	0.64	0.64	0.64
		NM	0.68	0.46	0.61	0.50	0.71	0.46	0.57
		ENM	0.61	0.68	0.71	0.79	0.79	0.68	0.79
local	high	E	0.86	0.93	0.96	0.93	0.93	1.00	1.00
		M	0.71	0.86	0.82	0.79	0.82	0.89	0.82
		N	0.50	0.64	0.61	0.61	0.64	0.71	0.79
		EM	0.89	0.96	1.00	1.00	1.00	0.96	1.00
		EN	0.93	0.93	0.96	0.96	1.00	0.93	1.00
		NM	0.79	0.82	0.86	0.89	0.82	0.82	0.89
		ENM	0.89	0.96	1.00	0.96	1.00	1.00	1.00
	low	E	0.68	0.64	0.75	0.75	0.82	0.86	0.93
		M	0.61	0.75	0.68	0.54	0.61	0.86	0.68
		N	0.61	0.50	0.29	0.57	0.39	0.71	0.46
		EM	0.75	0.93	0.86	0.93	0.89	1.00	0.96
		EN	0.68	0.68	0.71	0.82	0.89	0.86	0.86
		NM	0.64	0.71	0.68	0.79	0.75	0.79	0.75
		ENM	0.93	0.68	0.89	0.96	0.93	0.96	0.96

Table 4-11: Mean RT data (in ms) of correct trials for all experimental conditions. The PT conditions range from 32 ms (PT32) to 94 ms (PT94).

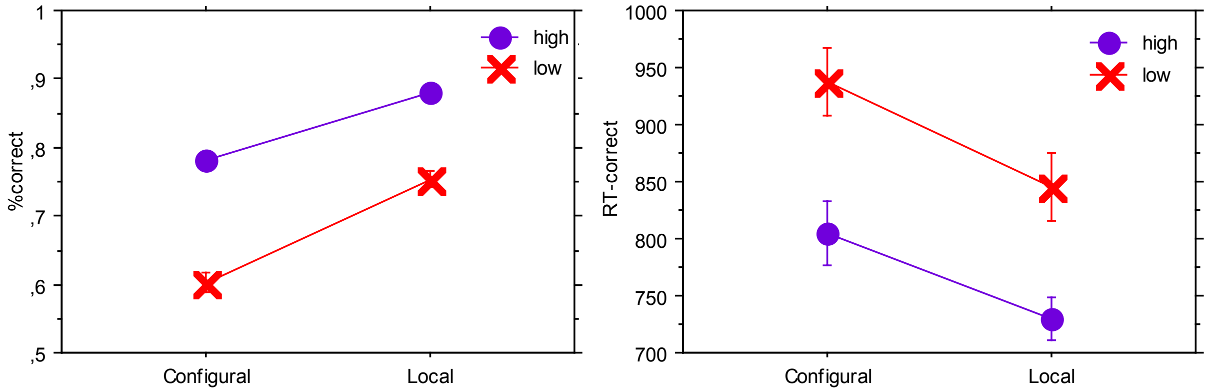
CLASS	SALIENCE	FEATURE	PT32	PT42	PT53	PT63	PT74	PT84	PT94
configural	high	E	983.75	746.75	819.32	643.06	890.83	650.42	670.15
		M	874.86	842.19	748.24	729.52	676.71	722.50	737.30
		N	874.40	807.18	732.20	879.40	883.35	732.53	783.44
		EM	750.57	784.72	724.36	754.33	683.96	636.12	645.04
		EN	905.71	714.05	728.17	694.46	734.40	644.39	782.63
		NM	822.68	845.10	1777.00	759.50	781.47	780.55	856.27
		ENM	886.00	738.28	719.42	636.50	756.70	694.24	712.11
	low	E	1112.80	847.92	808.53	885.59	853.55	941.00	1001.81
		M	952.92	777.62	1044.56	845.50	822.33	809.50	781.29
		N	1126.29	954.20	915.17	910.33	891.53	865.43	1011.27
		EM	767.00	822.96	782.78	841.00	850.17	874.89	704.27
		EN	1062.20	855.21	771.75	1050.06	785.33	709.17	652.39
		NM	872.32	816.23	1258.94	758.71	1078.55	964.31	953.19
		ENM	1124.24	890.47	785.40	765.00	953.59	713.68	824.14
local	high	E	725.33	647.77	673.44	704.77	674.54	728.68	700.61
		M	654.15	811.29	891.39	846.36	671.83	673.48	759.04
		N	643.36	680.67	720.77	680.24	815.44	761.65	864.77
		EM	679.24	692.30	612.89	623.79	582.96	661.00	648.89
		EN	784.42	677.58	603.00	662.44	701.50	637.15	651.11
		NM	792.96	814.44	663.71	670.96	621.48	738.83	1022.48
		ENM	689.16	643.82	699.71	644.74	635.93	603.07	726.68
	low	E	971.11	891.28	726.00	717.86	645.09	694.33	730.35
		M	844.18	1381.00	822.95	1111.60	816.59	829.17	683.68
		N	959.88	933.64	677.13	863.00	1046.27	964.35	1157.31
		EM	887.57	686.77	850.00	712.65	635.76	774.54	725.04
		EN	909.90	765.68	969.05	769.13	743.20	845.04	756.33
		NM	875.28	895.15	786.11	912.09	962.05	682.32	763.10
		ENM	879.42	673.05	695.40	668.15	634.77	682.04	702.33

Hyp.1–Salience

In order to test Hyp.1, the percentage correct data were submitted to a repeated measurement ANOVA with CLASS (configural vs. local), SALIENCE (high vs. low), FEATURE (E, N, M, EN, EM, NM, ENM) and PT as within-subjects factors (see Table 8-19 in the appendix for details). SALIENCE was found to be significant ($F_{1,27}$=77.3; P<.0001): highly salient parts were detected much better than their low salience parts (.83 vs. .68). Thus, Hyp.1 was confirmed.

In addition, a second repeated measures ANOVA was calculated with RT data for all correct trials. Analyzing the RTs will reveal whether highly salient faces are not only recognized better but are also recognized *faster* than their low salience counterparts. The design of the independent factors was reduced to three within factors (CLASS, SALIENCE, and FEATURE)[45]. Again, there was a significant SALIENCE effect: Highly salient features were detected much faster (767 ms vs. 891 ms). See Figure 4-18 for both salience effects, here combined with the CLASS factor.

Figure 4-18: Effects of the factor SALIENCE: highly salient features are recognized more accurately *and* faster than features of low salience (error bars are *SE*s). Additionally, an obvious influence of CLASS (local vs. configural) emerged. On the left, the percentage correct rate is shown (from .5 to 1.0); on the right, the RT for correct responses are shown from 700 ms to 1000 ms.



These results are again a validation for the contention that the distinctiveness data of the prestudies were operationalized in a psychologically relevant way. The salience data have a direct predictive power on the percent correctness rate and the RT pattern (analyses below). This could be interpreted in several ways:

First, according to Massaro (1989), salient stimuli are perceived better and faster due to their higher neural processing and their more efficient processing.

Second, Bachmann (1989) showed that a more intense first stimulus will dominate perception. In the case of a target face of low salience, the first stimulus, which is of medium salience, will therefore be dominant. The consequence is a dominance of the first picture, which lowers the recognition rate (of the second one).

Third, in borrowing from the line of reasoning proposed by Smedslund, the finding that highly salient features are recognized better and faster is a pseudo-empirical finding, because

---

[45] The ANOVA for RT-data (for correct trials) had to be reduced to three independent factors, because some cell combinations, especially those for very short PTs were empty, since only correct responses were included and the correctness of data was rather low for those PTs due to the difficulty of task. This also caused a decrease in degrees of freedom (*df*s) (subject-*df* of 21 vs. 27).

the terms *salience* or *distinctiveness* intrinsically contain the assumption that salient features *are* of a particular quality (Smedslund, 1979). Is a salient feature imaginable that is *not* salient?

Hyp.2–Feature dominance

The hypothesis about *Feature dominance* is based on an analysis of accuracy data for the single features (E, N, M). It is hypothesized that changes to some face areas are better detectable, namely that alterations of the eyes region are very well recognizable. In order to test Hyp.2, the percentage correct of the single features only was analyzed. This was done to investigate whether there is an advantage for single features over other single features. Therefore, a repeated measures ANOVA with the within factors CLASS, SALIENCE, SINGLEFEATURE (E, N, M) and PT (see Table 8-21 in the appendix for details) was conducted. The factor SINGLEFEATURE was significant ($F_{2,54}$=21.60; P<.0001). A conservative Scheffé post hoc test revealed that the differences of the possible pairs were significant (Ps<.0215; see details in Table 8-22 in the appendix). There was a clear trend that changed eyes were detected best (.77), followed by the mouth (.67), and the most difficult detection of the nose (.54). Given a base rate of 50% of same trial, the alteration of the nose was poorly detected (one-tailed T-test with a hypothesized mean of .5: T_{27}=1.10; *n.s.*). A similar result was reported by Haig (1986). The order of features is quite the same when the data is further split by factor SALIENCE with a spread between the percentages correct of .15; on the grounds of a general advantage of highly salient features, this spread will be called *salience advantage effect* (see Table 4-12 for more details). This kind of salience effect is well known in the face recognition literature: Both distinctive and attractive faces are recognized better than typical ones (Davies et al., 1977, for a review), as was already supported statistically by the ANOVA reported above.

Table 4-12: Hit rates for different pairs for the main features, split up by SALIENCE. The last row shows the difference of the hit rate between versions of high and low salience.

%correct (different)	Eyes	Mouth	Nose
High salience	.85 (.36)	.74 (.44)	.62 (.49)
Low salience	.70 (.46)	.60 (.49)	.47 (.50)
Salience advantage effect	*.15*	*.14*	*.15*

For all these data based on the variable of saliency it should be considered, however, that there does not exist any reliable measure. The distinctiveness scale for the eyes is not comparable with that of the nose or the mouth (cf. Young et al., 1985). It is only possible to distinguish between high and low distinctiveness objects *within* one scaling (e.g., a high distinctiveness nose vs. a low distinctiveness nose). This is not the case when the distinctiveness of *different* features is compared. In other words, it is not possible to ascertain whether differences in one scaling are transferable to another. What does it mean to compare a nose of high distinctiveness with a mouth of high distinctiveness?

When the percentage correct data are split up by salience, it can be demonstrated that highly salient noses can be detected, although the recognition performance is still near the base rate (.62; T_{27}=2.61; P=.0145).

All these tests were concerned with single features. However, is the assumption of the nose as a feature with nearly no benefit for the detection of changes also valid for combined feature changes? If it is the case that a changed nose cannot be detected at all, then an additional alteration of the nose should not be advantageous for recognition. For example, a simultaneous change of eyes and nose (EN) should not give a recognition advantage to a single changed eyes region (E). The same should be the case for NM against M. Moreover, by comparing E with EN and M with NM, there must be a difference between these pairs in respect to hit rates, if it is assumed that there is an indirect influence of the 'weak feature' *nose* when presented in combination with other changes. This would fit in a holistic processing theory, where not single parts on their own have an impact, but rather the sum of all parts. In Table 4-13 the percentage correct data of the two tested pairs (E vs. EN and M vs. NM) are shown with the results of referring one-tailed *T*-tests.

Table 4-13: Accuracy data used to test whether an additionally changed nose is advantageous for recognition. The table is split by salience. The *T*-column indicates the *T*-values of a one-tailed *T*-tests with *df*=28. The *P*-column gives the referring *P*-values of these *T*-tests.

Salience	E	EN	T_{27} (E,EN)	P	M	NM	T_{27} (M,NM)	P
high and low	.78	.81	1.49	.0738. n.s.	.67	.72	2.54	.0085
high	.85	.91	3.34	.0012	.74	.80	2.46	.0104
low	.70	.71	.14	.4437. n.s.	.60	.65	1.27	.1075. n.s.

The difference of E and EN was not found to be significant (two-tailed *T*-test: T_{27}=1.49, n.s.). The difference between M and NM, on the other hand, was significant (T_{27}=2.54, P=.0085). By further analyzing the data, it was obvious that only the highly salient faces were responsible for any additive nose effect. By only calculating highly salient faces, there emerged a difference between E and EN as well as between M and NM (see for *T*-values Table 4-13). However, there were no effects for faces of low salience (Ts_{27}<1.27, n.s.). This lends support to a holistic processing strategy: For highly salient faces, an additionally changed nose in both cases had a positive influence on the detectability of changes.

Hyp.3–Feature quantity

Are the effects referring to additionally changed noses transferable to other features? Hyp.3 assumes that the more features are changed, the better the recognition rate will be. Table 4-14 shows the accuracy data ordered by faces with one, two, and three changed regions.

Table 4-14: Accuracy data ordered by faces with one, two, and three changed regions. The last column shows the *SD*s.

Number of features	Correctness	*SD*
1 (E, N, M)	.663	.31
2 (EM, EN, NM)	.804	.27
3 (ENM)	.878	.33

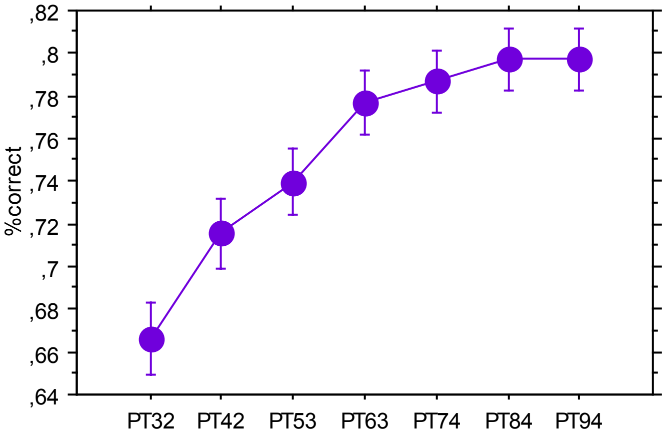
Hyp.3 was tested with a repeated measurement ANOVA with CLASS, SALIENCE, FEATUREQUANTITY (number of features that differ between basic- and target-face: 1:E,N,M; 2:EN,EM,NM or 3: ENM) and PT as within factors (see details in Table 8-23 in the appendix). The most interesting factor is FEATUREQUANTITY. Testing this factor for significance is a direct test of whether a combination of feature changes compared with single features has a positive influence on the recognition rates. The factor FEATUREQUANTITY was indeed significant ($F_{2,54}$=67.7; P<.0001). Scheffé post-hoc tests revealed significant differences between all observed pairs (*mean diff*s>.073; *P*s<.0012). The predicted order was confirmed: the recognition was best when participants were confronted with three changed locations and it was worst if only one feature had been changed. Again, as with the result of single features, there was no interaction between FEATUREQUANTITY and CLASS, which reveals that the kind of CLASS has no qualitative influence on the recognition of features when all PTs are aggregated. By taking into account the PTs, this result is different. A strong trend for the interaction of CLASS, FEATUREQUANTITY and PT ($F_{12,324}$=1.73; P=.0589, n.s.) reveals that PT affects the interaction between CLASS and FEATUREQUANTITY.

What does this mean? With such a three-way interaction, it can be assumed that the detection of changes develops differently over the first 94 ms, depending on the quality of these changes (configural vs. local). Thus, Hyp.3 is generally supported. However, by varying PTs, dissociate effects of FEATUREQUANTITY with CLASS emerged. This is a hint that the processing of different feature classes does not run fully parallel. This dissociation will be further investigated in the discussion of Hyp.6.

Hyp.4–Validation of PT

Before testing specific hypotheses about the influence of PT on the recognition of the different face classes, it has to be validated whether PT as such has an influence on recognition accuracy (Hyp.4). This analysis will reveal whether the range of PTs is set well. To test this, the significant factor PT of the ANOVA calculated for testing Hyp.1 ($F_{6,162}$=10.41; P<.0001) was analyzed further (see full ANOVA data in Table 8-19 in the appendix). As indicated in Figure 4-19, there was a clear monotonous increase of recognition performance from the starting point of 32 ms to 94 ms (with equal scores at 84 ms and 94 ms and therefore not in a *strict* monotonous fashion).

Figure 4-19: Validation of the PT setting. Increases in performance with increasing PT are shown on the recognition rate (see data in Table 8-24 in the appendix).



The low rate at 32 ms of only .666 (see Figure 4-19) demonstrates that it is very difficult to process faces after such a brief presentation. Nevertheless, after 84 ms the recognition performance reaches 80%. Thus, the manipulation range from 32 ms to 94 ms was well chosen. Consequently, it is worth to investigate different PT-effects on other variables. This validates Hyp.4.

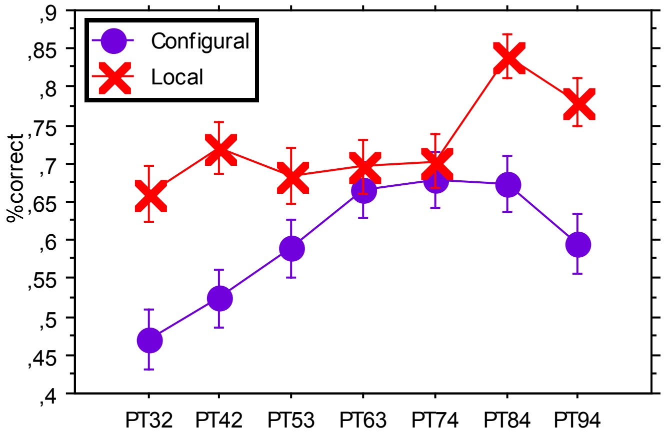
Hyp.5–Rapid identification

In order to test Hyp.5, it must be checked whether these recognition rates are above the general guessing rate of 50%. The data revealed that even after 32 ms it is already possible—although at a low level of only .666—to recognize changes in faces. A T-test demonstrated that even with such a short PT, the recognition rate is *above* the chance level (two-tailed T-test with T=8.56, P<.0001). This gives support for Hyp.5. Moreover, it adds evidences to Bachmann's (1991) finding that very early face percepts are available within only 100 ms.

Hyp.6–Configural and local feature processing

It was predicted that the different face classes (configural vs. local) might not be processed simultaneously. To test the hypothesis of a processing-order, with locally changed items being recognized earlier than configural ones (Hyp.6), it is useful to look at the recognition data again. As illustrated by the data for the single features only (Figure 4-20), there is an interaction between CLASS and PT.

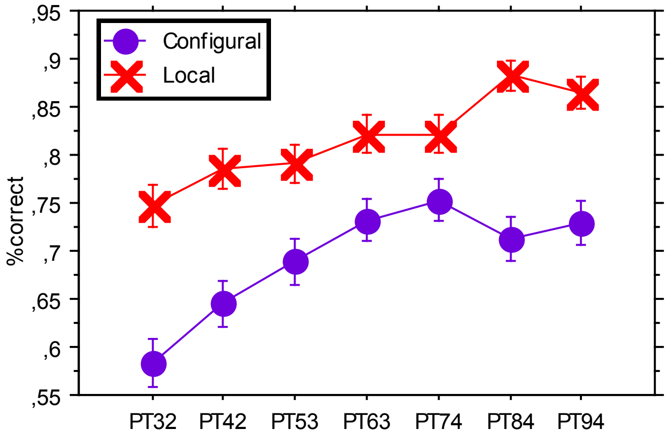
Figure 4-20: Relationship between CLASS and PT for different pairs (only single features are included). See data in Table 8-25 in the appendix.



A repeated measurement ANOVA, which was already described for testing Hyp.2, revealed a significant interaction of CLASS and PT ($F_{6,162}$=3.92; *P*=.0011). The change detection process looks as follows: local detections are possible very early and until about 80 ms in a constant, but modest quantity (.66≤rate≤.72). After about 80 ms, the rate rises over .78. The development of the detection of configural changes on the other hand benefits continuously from a longer PT until about 80 ms (.47≤rate≤.67). However, after remaining stable for about 30 ms at a level of about .67, the configural recognition at 94 ms drops again to a rate of .60.

A similar result—only as a trend, however—was found by including every feature combination in the analysis ($F_{6,162}$=2.13; *P*=.0523; see data in Table 8-26 in the appendix). The most important difference between both analyses is the more harmonious curves' quality when including all data, probably due to the more data points aggregated for this plot (see Figure 4-21).

Figure 4-21: Relationship between CLASS and PT for different pairs (all features and their possible combinations included). See data in Table 8-26 in the appendix.



Additionally, the analyses of the RT data (for correct trials), reported above for testing hypothesis Hyp.1 (see full ANOVA results in Table 8-20 in the appendix), revealed a significant factor CLASS ($F_{1,21}$=16.49; *P*=.0006). Local features were recognized sooner than configural ones (788 ms vs. 870 ms; see illustration in Figure 4-18).

Despite the pre-experimental non-difference between configural and local manipulation shown by distinctiveness ratings in Pre-Study 2, significant differences were obtained here. What might be the reason for this discrepancy? There are several possible explanations.

On the one hand, it could be possible that distinctiveness ratings are highly correlated with recognition performance, but are not based on the same cognitive processing. It is conceivable that recognition processes are more sensitive to differences in saliency than are ratings.

On the other hand, by looking at Figure 4-21, it is likely that only the linearity assumption of the ANOVA testing model is responsible for the statistical difference between configural and local data. As discussed in several papers by Geoffrey R. Loftus (e.g, Loftus, 1989, 1995, 1996; Loftus & Loftus, 1987; Loftus & Masson, 1994), it is only a commonly used but not obligatory procedure to test hypotheses in experimental psychology with a *linear* model. Alternative interpretations are possible. For instance, it is conceivable that the configural recognition process only begins later, but reaches the same performance level after a while. Thus, the curves are probably only shifted, with the starting point of configural recognition appearing approximately 30 ms later. The experimental design provides only seven data points for analyzing each curve. Consequently, it is speculative to base this conclusion on such sparse data, because there is nearly no overlapping recognition performance of the two CLASS-curves. However, what seems clear is that in a linear model as well as in an alternative model, configural information is processed later than local information. The precedence of local feature analysis can be interpreted in favor of Biederman's *RBC theory* (Biederman, 1987). Biederman postulated a hierarchical set of processing with an accentuation on early feature-based recognition. The combination of several components defines *geons*, which are rudimentary three-dimensional basic features.

Nevertheless, it cannot be ruled out that global structures like the outline or a low-spatial image of the face are processed even earlier than local information. The distinctiveness data of Pre-Study 1a tentatively favor this assumption. Besides the ratings of the inner features, the distinctiveness of the hair region was evaluated. The relatively high distinctiveness rating for the hair area (see Table 4-3) revealed that the hair region might be important for the face recognition process. Other evidences for an important influence of the hair area on *recognition performance* was found in many studies (see Valentine, 1999). As the largest representative of facial global structures, it would be an important cue for the processing of faces. The early processing of this cue is only speculative, because the experimental design did not control for the amount of global information. Therefore, such an early global recognition processing is not implausible, but also not empirically testable with the data of Experiment 1.

Hyp.7–Feature processing

The last hypothesis investigates the microgenetic processing of faces. According to the microgenetic processing assumption, a face is not processed at once, but in a succeeding sequence of single analyses.

In this section, specific processing models have been tested. The aim was not to investigate whether *there is any* microgenesis occurring, but *what* the exact nature of the facial processing is and which feature-dependencies exist. Therefore, it is the most essential part of hypotheses testing of the whole experiment.

Recognition performance for single features

In order to understand the time-course of feature processing, it is important to analyze the development of recognition performance within the microgenesis for every single feature. From these data it can be gleaned from what time on a feature can be processed. In the later part of this section, all combinations of features are also included in another analysis. In combination with a newly introduced measure called *WOM*, models that are more complex will be tested.

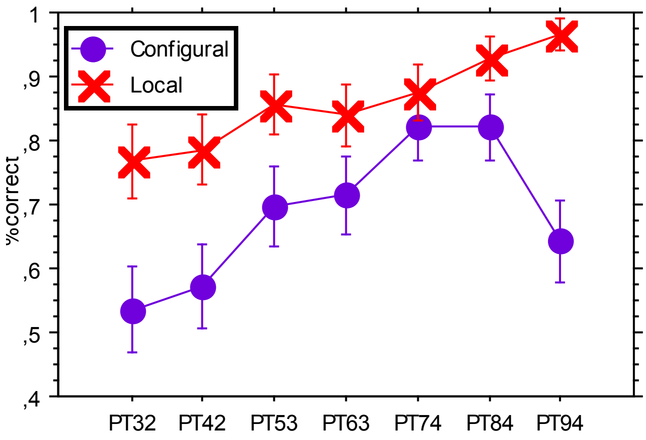
Table 4-15 shows the data for each single feature for every CLASS * PT combination. All features have to be analyzed for that PT for which the recognition is yet above the chance rate.

Table 4-15: Recognition data (percentage correct) for the single features E, N, and M for all CLASS * PT.

CLASS	PT	Eyes	*SD*	Nose	*SD*	Mouth	*SD*
Configural	32 ms	0.536	(0.50)	0.393	(0.49)	0.482	(0.50)
	42 ms	0.571	(0.50)	0.482	(0.50)	0.518	(0.50)
	53 ms	0.696	(0.46)	0.482	(0.50)	0.589	(0.50)
	63 ms	0.714	(0.46)	0.625	(0.49)	0.661	(0.48)
	74 ms	0.821	(0.39)	0.571	(0.50)	0.643	(0.48)
	84 ms	0.821	(0.39)	0.554	(0.50)	0.643	(0.48)
	94 ms	0.643	(0.48)	0.482	(0.50)	0.661	(0.48)
Local	32 ms	0.768	(0.43)	0.554	(0.50)	0.661	(0.48)
	42 ms	0.786	(0.41)	0.571	(0.50)	0.804	(0.40)
	53 ms	0.857	(0.35)	0.446	(0.50)	0.750	(0.44)
	63 ms	0.839	(0.37)	0.589	(0.50)	0.661	(0.48)
	73 ms	0.875	(0.33)	0.518	(0.50)	0.714	(0.46)
	84 ms	0.929	(0.26)	0.714	(0.46)	0.875	(0.33)
	94 ms	0.964	(0.19)	0.625	(0.49)	0.750	(0.44)

The development of the percentage correct for ‘eyes-changes *only*’ is very similar to the recognition rate curves for all PTs shown above. Analogous to those data, there was an obvious difference between local and configural faces. Additionally, there was again an increase in performance for longer PTs[46] (see Figure 4-22).

Figure 4-22: Relationship between CLASS and PT for ‘EYES only’ in terms of recognition rates.



---

[46] With the exception of configural recognition for PT=94 ms.

Revealing starting points of processes with α-adjusted T-tests

For generating a simple microgenetic model, it is not very important to base the model on exact performance rates. The essential data for such a model are the *starting* points of single processes. Therefore, only those recognition rates of Table 4-15 are essential which are *above* chance level (.5). For every PT and CLASS a two-tailed *T*-test was calculated and compared with the chance probability. Due to parallel tests, the α–level had to be adjusted to .0073 according to the Bonferroni adjustment rule (Kirk, 1982; Winer, 1971). The α-level for each *T*-test had to be lowered to $\alpha' = 1-(1-\alpha)^{(1/7)}$. This routine brings the overall α'–level back to the desired α (.05). By doing so, this technique revealed that configurally changed noses and mouths did *not* differ from the base rate. It did not matter whether a long or only a short PT was given, the recognition performance was equal to chance. This means that within the time range of the first 94 ms, a configural changing of the area around the mouth and the nose could not be detected. The only configurally changed item which was recognizable within this time period was the eyes region.

Nevertheless, the Bonferroni adjustment method used here might fail to show important differences due to its *conservative* criterion. Methods like the Bonferroni adjustment that are conservative about committing false-positive errors are consequently rather liberal about committing false-negative errors, i.e. accepting the null-hypothesis although there is a significant difference. The problem to miss real differences in favor of falsely accepting null hypotheses is well known (Bland & Altman, 1995), but not yet solved (Perneger, 1998; Thomas et al., 1985)[47].

Regression analyses for single features

A *simple linear regression* with PT as independent and percentage correct as dependent variable could provide evidence for a detectable mouth within the first 94 ms. The *Pearson* correlation of R=.829 indicates a strong dependence of recognition performance on PT. The regression was found to be significant by a factorial ANOVA with percentage correct as dependent and PT as independent variable ($F_{1,6}$=10.97; P=.0212). Moreover, the regression equation (for PTs measured in seconds) had a positive gradient and an intercept above .5 (see Equation 4-3).

Equation 4-3: Term for the *simple linear regression* with recognition rate (correct) of mouth as dependent variable.

$$correct_{mouth} = .541 + 2.086 \cdot PT; R^2 = .687$$

The gradient is not as high as for the eyes-regression (see Equation 4-4). Nevertheless, in combination with an intercept of .541 it is obvious that the detection of mouth developed within the first 94 ms of presentation.

Equation 4-4: Term for the *simple linear regression* with recognition rate (correct) of eyes as dependent variable.

$$correct_{eyes} = .572 + 3.178 \cdot PT; R^2 = .754$$

---

[47] For instance, Thomas et al. (1985) preferred a methodical approach in which all associations in the data are reported, whether significant or not, followed by a ranking in order of priority for investigation using empirical *Bayes* techniques. Moreover, Feise (2002) emphasized that *P*-value adjustments are focused upon a universal and thus irrelevant null hypothesis (cf. Loftus, 1995, 1989).

The regression analysis of the recognition rate of the nose revealed similar results. However, with a lower intercept and a lower gradient it seems that the processing of the nose starts later than the processing of the mouth.

Equation 4-5: Term for the *simple linear regression* with recognition rate (correct) of nose as dependent variable.

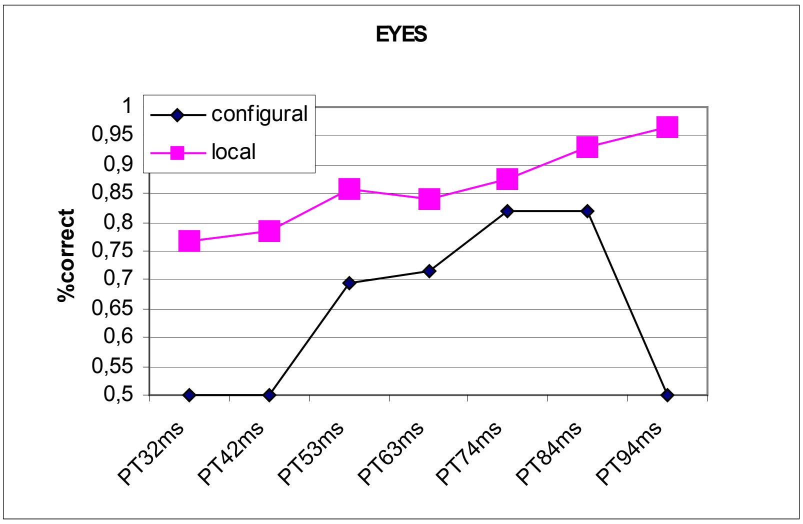
$$correct_{nose} = .427 + 1.841 \cdot PT; R^2 = .429$$

The different starting points of processing the eyes, the nose, and the mouth are in accord with the consideration of a shifted processing of configural vs. local features, discussed within the testing of Hyp.6.

Revealing starting points of processes with non α-adjusted T-tests

The trend of a detectable mouth area was also confirmed when the α-level for the *T*-tests described above was not adjusted. The nose again does not have the power to be detected within the given time range (with one exception under 84 ms). Nevertheless, the detection of the mouth is under these circumstances above the base rate for all PTs in the local version and after 63 ms in the configural version. To illustrate the non-α-level adjusted results, a diagram will be shown which combines inference statistical results with graded information about the height of recognition rate. When data were above chance (revealed by one-tailed *T*-tests), it will be shown as absolute. If it does not differ significantly from chance, it will be integrated as high as the chance level of .5. This kind of illustration is rather uncommon and not unproblematic, because a significant difference from chance can only indicate a *probability* to be the real value. Nevertheless, it is an intuitive illustration of *when* a process is going to start with indications of how strong this process is developing. Therefore, absolute values only indicate the strength of 'effective' processing[48]. Figure 4-23 shows that locally changed eyes are yet recognizable after only 32 ms with a rather high performance, whereas configural changes are recognized less well *and* later.

Figure 4-23: The single feature EYES with percentage correct in dependence of the PT. All non-significant differing rates are set to .5 (base rate) to make the development of the detection rate clear (used post-hoc tests are not α-adjusted).



---

[48] These analyses are based on the data set of recognition rates for the single features eyes, nose, and mouth shown in Table 4-15.

Nevertheless, Figure 4-23 reveals one problematic finding. There is an obvious drop of recognition rate for PT=94 ms. Why should a longer PT *decrease* the recognition quality? The raw data in Table 4-15 shows nothing conspicuous in the standard deviations, thus this effect seems to not be due to a measurement error. It rather seems that the devastation of the monotony is similar to the results of Bachmann (1991; 1987). He found that coarse identification processes lead to a decrease in performance when the exposure was increased. As explained by Bachmann, the availability of coarse-scale information, which is sufficient for good performance at shorter exposures, should still be there, but coarser information is no longer used at longer PTs. The same was found by fine-scale detailed information with just the opposite order of processing. In the example above, configural information would be the 'coarse information' in Bachmann's terms. This might explain the 94 ms drop, but is not able to describe the low or non-existent recognition performance for configural information at very short PTs. A general drop of performance with longer PTs was also found by Read, Vokey, and Hammersley (1990). They revealed that with high similarity between two faces, increased exposure duration always improved subsequent recognition. However, when the two representations of a person were judged to be low in similarity, increases in exposure duration served to reduce recognition performance. These results, although obtained by an experimental paradigm with much longer PTs, are important to demonstrate that an increase of PT not always yields better recognition! Nevertheless, it is statistically problematic to argue towards a dissociation of processes with only having one salient data point, which does not follow the monotonous trend!

The resulting pattern for the nose area is quite different from that of the eyes. The nose-change detection was above the chance level only once, namely in the local-84 ms condition. Figure 4-24 illustrates these findings.

Figure 4-24: The single feature NOSE with percentage correctness in dependence of the PT. All non-significant differing rates are set to .5 (base rate) to make the development of the detection rate clear (used post-hoc tests are not α-adjusted).

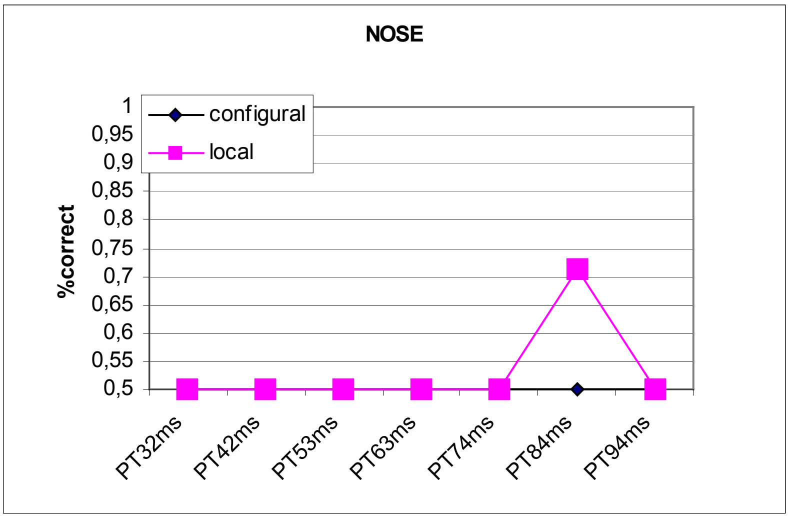
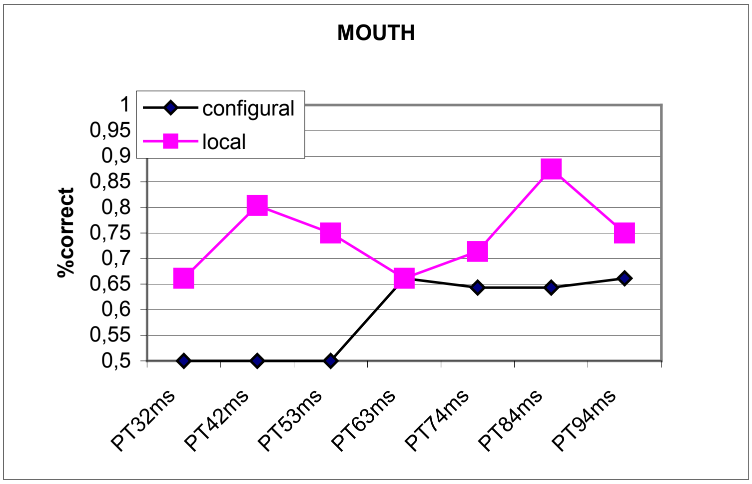


Figure 4-25 shows the recognition performance when only the mouth region was changed. Analogous to the findings for the eyes condition, the local versions were recognized sooner and better than their configural counterparts were. However, the configural data were generally on a lower level.

Figure 4-25: The single feature MOUTH with percentage correctness in dependence of the PT. All non-significant differing rates are set to .5 (base rate) to make the development of the detection rate clear (used post-hoc tests are not α-adjusted).

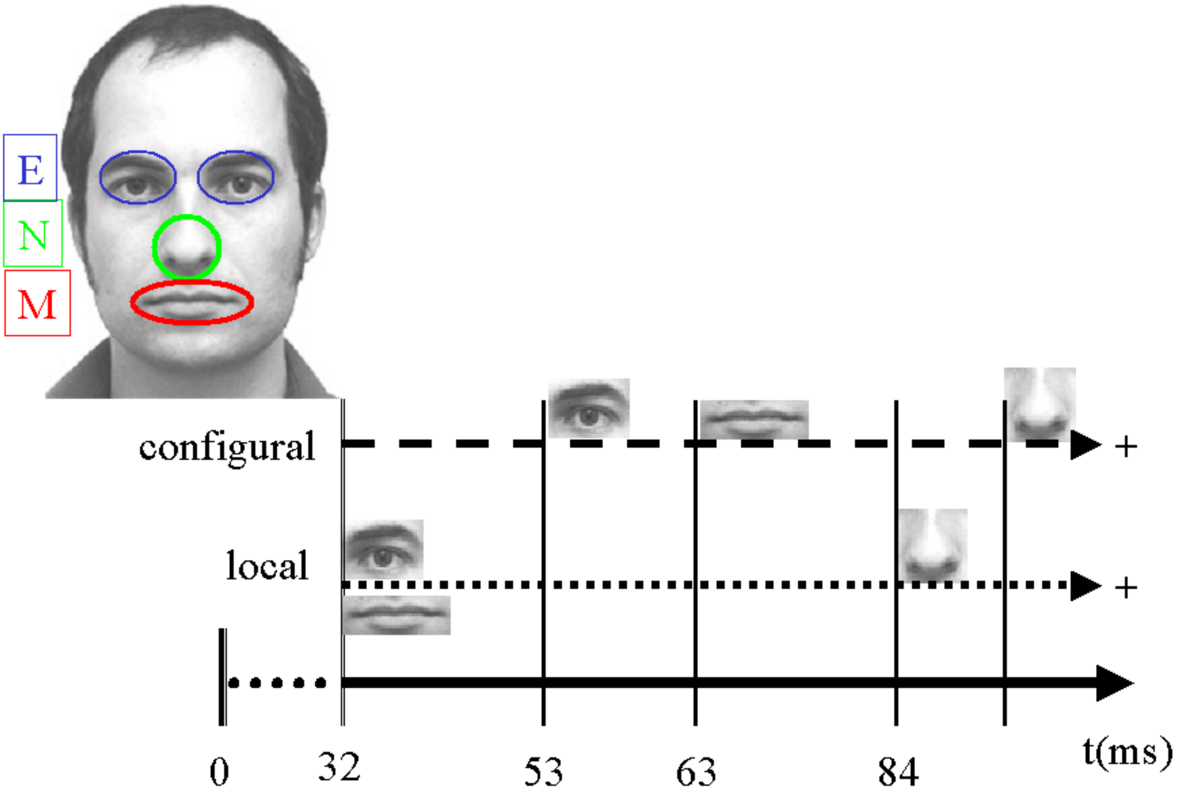


To summarize these results from the analyses presented here, the eyes region was the first inner feature of a face that was processed. This was particularly the case for locally changed eyes. It must not be forgotten that this and the following assumptions were based on the manipulation of only the eyes, nose, and mouth. Therefore, any other features like the chin or the forehead were not considered. The detection processes for configural changes already began at about 53 ms (see illustration for eyes region in Figure 4-23). The mouth was also a relatively salient part of the face and therefore has the power to be processed already after a short PT of 32 ms in the local condition. Analogous to the eyes processing, although a little bit later, configural changes were registered. The time needed for this was at least 63 ms (see Figure 4-25). However, the recognition performance of changed mouths clearly lagged behind the one of eyes: when only the nose area was altered, the recognition rate was worst. Changes of the nose could hardly be detected within the given 94 ms (see Figure 4-24), but the rise up of the hit rate for 84 ms under the local condition is maybe an indication that the processing began at this moment.

FACEREC-1: A model for face processing concerning the starting points of single processing

A possible model for detecting changes in the first 94 ms, which refers to the sum of these results, is FACEREC-1 (see Figure 4-26). In the FACEREC-1 model it is shown what kind of feature belonging to what CLASS (configural vs. local) will be detected at what time (in ms).

Figure 4-26: FACEREC-1: Model for the developmental stages of the recognition of a face. The first detection of a feature is indicated by an icon of this feature.



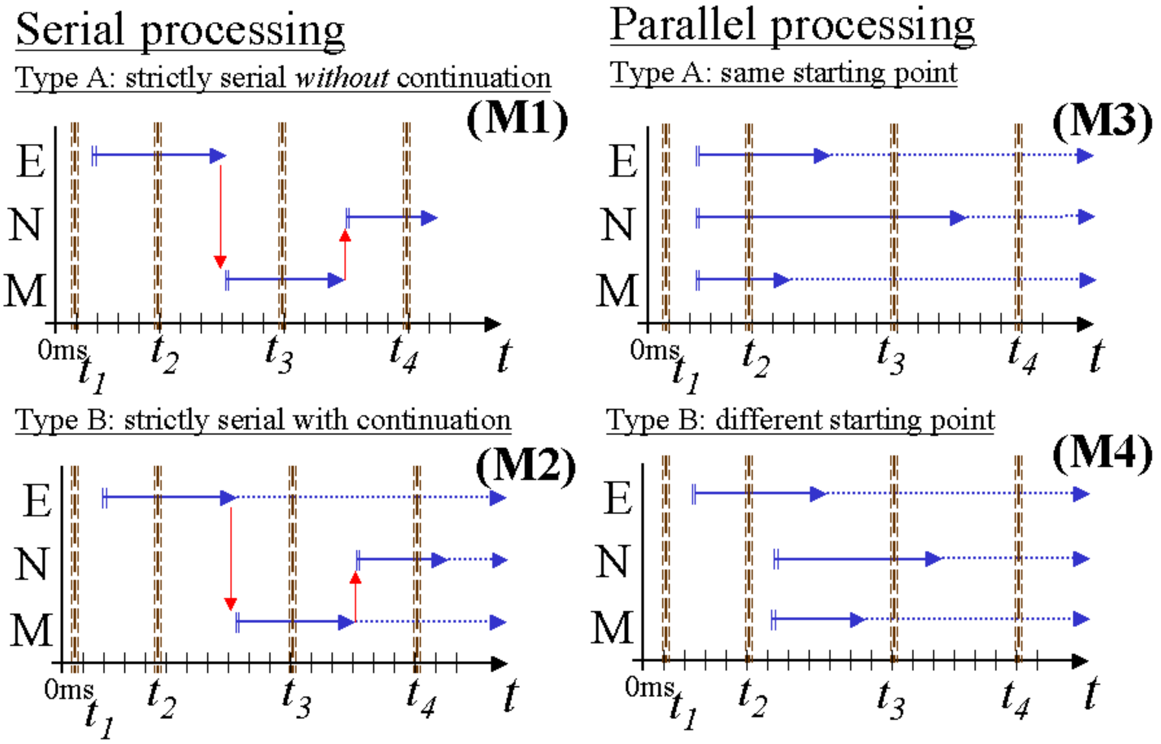
Such a model is in accord with the microgenetic account, especially in face processing (e.g., Sergent, 1986b; Watt, 1988), because the single features are not processed *at once* and synchronously, but in processing *stages*. The different importance of eyes, nose, and mouth was shown here in terms of face *processing* as prioritizing of the different features. Thus, the main revelation is that not all face regions are processed at the same time. There seems to be an order of feature processing and an order of the processing of different feature classes (with local features first and configural being analyzed later). It seems that there are at least two processes coming together in face recognition: an early detection process for local features and a detection of configurations that commences later. Within both of these modes, the eyes play a role of paramount importance. Changes to this region are very distinctive, which is in accord with earlier results (e.g., Haig, 1984; Hosie et al., 1988; Leder & Bruce, 2000b; Kemp et al., 1990). The specific operationalization of configural changes to eyes seems to give further advantage to the recognition of eyes. As Kemp et al. (1990) pointed out, subjects were more sensitive to detect changes of the eyes region when the eyes were displaced inwards or outwards than when there were vertical displacements. In Experiment 1 and Pre-Study 2 an inward shifting of eyes was engendered, but *not* a vertical displacement. Therefore, it might be advantageous to recognize eyes moved inward in comparison to vertically displaced noses or mouths. Nevertheless, the strong eyes dominance effect is *not* caused by replacing artifacts. Moreover, the replacement of eyes is much easier to realize and causes fewer graphical artifacts and substituting problems than, for example, exchanging a nose. By replacing a nose, it is difficult to adapt the skin color to the texture of the surrounding area. Despite this, the eyes have been recognized *sooner* and *better*.

Model FACEREC-1 represents the recognition data of all faces (with low as well as with high salience features). Thus, it cannot be excluded that particularly salient parts can be processed before the starting time points assumed by the model.

Testing recognition data of combined features

However, what can be the underlying mechanisms responsible for an ordering of feature processing? In Pre-Study 2 it was revealed that CLASS was not crucial for the evaluation of distinctiveness. Thus, it is plausible that a top-down process is involved to prioritize information. As section 3.2 and the introduction of this chapter described, there are several conceivable alternative processing models. The main models are based on either serial or parallel processing assumptions. To systematically and explicitly test such processing hypotheses, four common models and their specific assumptions about feature order and feature dependency will be presented. As are shown in Figure 4-27, in the following two paradigmatic serial processing models and two typical parallel ones will be tested.

Figure 4-27: Four different processing models. M1 and M2 with serial processing assumptions, M3 and M4 with parallel ones.



In Model M1, it is assumed that features are processed strictly sequentially with eyes first, then the mouth, and lastly the nose. This would be a typical assumption concerning the saliency-distribution (Koch & Ullman, 1985) but would be in contrast to a rigid *top-to-bottom* scanning strategy (see also Goldstein & Mackenberg, 1966; McKelvie, 1976; Davies et al., 1977), where an eyes-nose-mouth direction is proposed. The subsequent processing is only possible when the preceding one has already been completed.

Model M2 is similar but less constrained. For Model M2, it is assumed that the sub-processes of the recognition of single features do not have to be completely finished, before succeeding recognition processes of other features can start. Thus, the processing of the features is once again interdependent, but they are not synchronized by the end of the processing of the preceding feature.

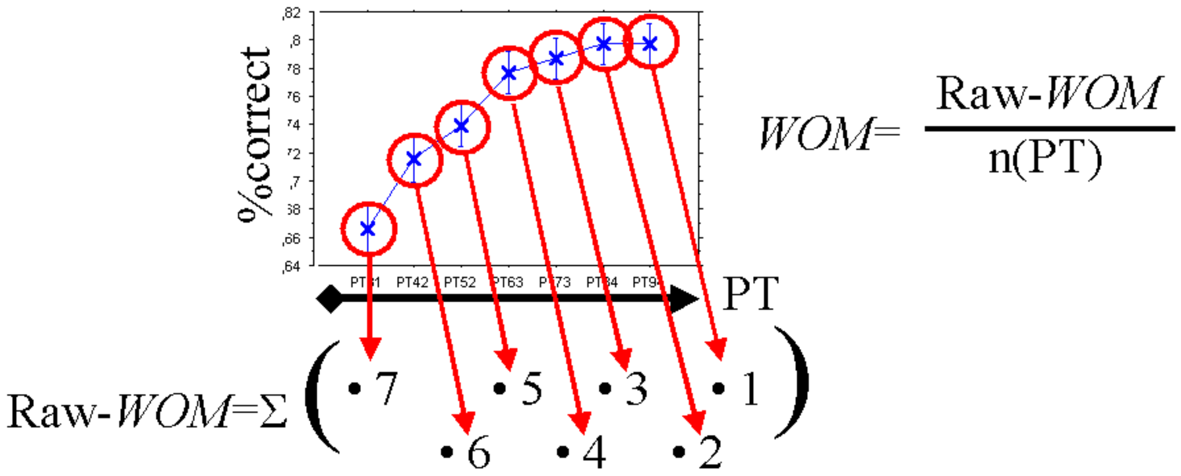
In contrast, Model M3 and M4 belong to a different class of processing models, because their key assumption is that the single recognition processes of features are *not* dependent upon each other, but are occurring simultaneously. Model M3 is the very strict version in which all processes start at the same moment.

Model M4, on the other hand, is a more liberal one with a non-synchronized commencement. Again, Model M4 assumes, as do M1 and M2, that the order of the starting of the single feature processes are guided by the saliency or the informative content of the facial structure.
In order to test these models and to compare them with each other, the recognition rates of every single feature *and* the combinations of these features for the different PTs have to be analyzed. This procedure is symbolized by the different *t*s (t_1, t_2, ..., t_4) in Figure 4-27. In the concrete example, the measuring of recognition performance at t_1 is not essential for model testing, because every model predicts that at such an early time no recognition is possible. However, discrimination between the models is already possible at t_2 (Model M3 versus the others), and so on. The integration of the recognition rates at all *t*s potentializes the falsification of several models.

Usage of WOM for testing specific processing models

Testing all model hypotheses against each other for *every* PT is complex and has not much power. Therefore, an integrative measure is needed. This measure combines the data of all *PT*s, without ignoring the extraordinary role of early recognition achievement. A pure arithmetic averaging would make *no* difference between early and late visual processes and would thus be inappropriate for solving this problem. A special measure would be needed, which takes into account whether a feature is recognized very early, because only such a variable can be used to test different temporal hypotheses.
An ideal tool for this requirement is a *weighted order measure*, in the following briefly called *WOM*. The *WOM* is a combination of weighted percentage correct rates, with multiplying PT32-rates by 7, PT42 by 6, etc. To get a normalized measure, this 'raw-*WOM'* will be divided by the number of PTs, here by n=7 (see Figure 4-28 for an illustration).

Figure 4-28: Calculation of *WOM*. First, the correctness rate of PT32 is multiplied by 7, of PT42 by 6, of PT53 by 5, etc. Second, the sum of these products is standardized by the number of PTs, in this case by *n*=7.



With the assumption of a monotonous increase of correctness with longer PTs, *WOM* inherently provides information about the *earliness* of a recognition process. This can be easily demonstrated by a simple theoretical example. Even if an item is recognized at a very high level but at a late processing stage—that is, at a long PT—then due to the low weighting of the late correctness rates this cannot compensate for the lacking early recognition.

Using *WOM*, specific model assumptions can be tested. These models refer to the four versions described in Figure 4-16. For each model a short description of its processing character is given. Supplementally, empirical predictions for *WOM* rates are given to test the models empirically.

Model M1

A strictly serial processing account (M1) would make the following two general predictions. First, there is a processing sequence of features. It is plausible that features are processed in the order of their information content and their social relevance. Thus the following order is assumed: the eyes first, then the mouth, and lastly the nose (cf. section 3.2.1). Second, because a progression of feature processing is only possible when the preceding feature is fully identified/recognized, there cannot be any advantage of multiple changed features over the single feature of the highest priority. This means, that an assumed top priority eyes region is recognized as fast as a combination of eyes and any other feature(s). The same would be the case for the mouth region of medium importance, as compared to the nose. Referring to *WOM*, the following empirical hypotheses are evident[49].

- *WOM*(E) > *WOM*(M) > *WOM*(N);
- *WOM*(E) = *WOM*(EM); *WOM*(E) = *WOM*(EN); *WOM*(E) = *WOM*(ENM);
- *WOM*(M) = *WOM*(NM);

Model M2

For Model M2, the order of features would be the same, but now, in spite of the not completely needed processing of preceding features, a combination of specific features should be advantageous:

- *WOM*(E) > *WOM*(M) > *WOM*(N);
- *WOM*(E) < *WOM*(EM); *WOM*(E) < *WOM*(EN); *WOM*(E) < *WOM*(ENM);
- *WOM*(M) < *WOM*(NM);

Model M3

A completely different set of predictions would emerge from a parallel account with restricted starting points. For Model M3 there are the following assumptions:

- *WOM*(E) = *WOM*(M) = *WOM*(N);
- *WOM*(E) < *WOM*(EM); *WOM*(E) < *WOM*(EN); *WOM*(E) < *WOM*(ENM);
- *WOM*(M) < *WOM*(NM);

Model M4

Model M4 is structurally different from the serial accounts, but it is very similar to M2 concerning the assumptions for *WOM*, because it is difficult to differentiate between shifted starting points (M4) and serial dependence that is *not* synchronized by the ending of a featural processing (M2). The evident difference between M3 and M4 is in the order of features. While there is a parallel starting in M3, there is a priority for E in Model M4. To make M4 distinguishable from M2, it is further assumed that the lower priority M and N are processed in parallel (with the same starting point). To sum up these assumptions, there is an order of features (E first, then M and N) and due to the parallel processing character, the more features

---

[49] There are, of course, many other empirical hypotheses testable for the single models, but only those assumptions which held contrast-information by assuming relations contrasted by other model-predictions are included here. Otherwise, it could not be fully compared which model fits best to the empirical data.

are altered, the better is the recognition. This yields the following empirical assumptions for Model M4:

- *WOM*(E) > *WOM*(M) = *WOM*(N);
- *WOM*(E) < *WOM*(EM); *WOM*(E) < *WOM*(EN); *WOM*(E) < *WOM*(ENM);
- *WOM*(M) < *WOM*(NM);

Empirical testing of specific processing models

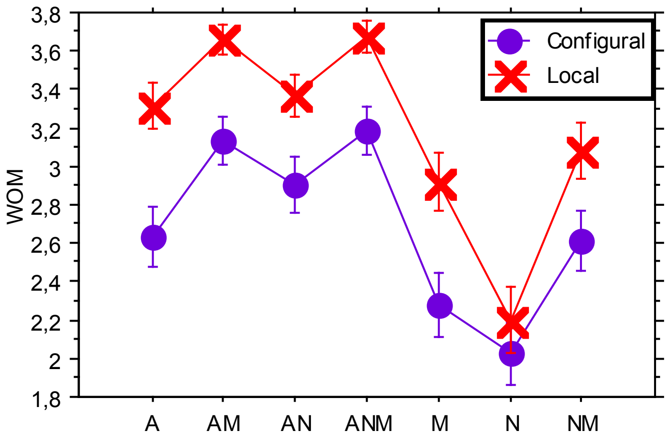
In section 3.2 on face processing and possible processing models, the distinctions between analytical and holistic, and between serial vs. parallel processing were introduced. There is an ongoing debate on how faces are processed. However, only rarely is the quality of information that has to be recognized taken into account for testing these models. For example, configural and local features might be processed in a different way. By confounding configural and local manipulations, the true nature of processing could be concealed. This might be the reason for the existence of a plethora of contradictory data confirming several distinct processing models. In the following analyses, configural and local versions were tested independently for the above-described models. The aim was to find dissociate processing for both classes to explain the ambivalent findings of several face recognition experiments. A reason for ambivalent evidence for different processing types might be a confounding of configural and local manipulation. In Experiment 1, this confounding problem was solved by splitting facial manipulations into configural *or* local ones. The raw data of *WOM* values for analyzing the underlying processing is shown in Table 4-16.

Table 4-16: *WOM* data for all features of both face classes. In the first data column, both saliency levels (high and low) are aggregated. The next two columns show the *WOM* data for high and low salience manipulations only. The *SD*s can be found in brackets.

CLASS	FEATURE	High & Low salience	High salience	Low salience
Configural	E	2.633 (1.14)	2.998 (1.04)	2.269 (1.14)
	EM	3.131 (.96)	3.538 (.67)	2.724 (1.04)
	EN	2.903 (1.09)	3.287 (.89)	2.519 (1.14)
	ENM	3.183 (.94)	3.556 (.72)	2.811 (1.00)
	M	2.280 (1.23)	2.497 (1.28)	2.064 (1.15)
	N	2.025 (1.22)	2.391 (1.23)	1.659 (1.10)
	NM	2.616 (1.17)	2.904 (1.06)	2.328 (1.23)
Local	E	3.314 (.91)	3.714 (.63)	2.915 (.98)
	EM	3.660 (.62)	3.843 (.46)	3.477 (.71)
	EN	3.371 (.82)	3.782 (.44)	2.959 (.91)
	ENM	3.674 (.64)	3.842 (.44)	3.506 (.77)
	M	2.919 (1.16)	3.203 (1.23)	2.635 (1.03)
	N	2.197 (1.29)	2.409 (1.48)	1.985 (1.04)
	NM	3.083 (1.07)	3.300 (1.01)	2.866 (1.11)

To test which of the above-described models best fits these data, specific differences between pairs of faces must be tested. For example, Model M1 assumes that *WOM*(M)=*WOM*(NM), whereas all other models predict *WOM*(N)<*WOM*(NM). An illustration of the *WOM* data in Figure 4-29 shows that concerning this single assumption, Model M1 best fits the serial data. In contrast, the other models (M2, M3, M4) are compatible with the configural data set.

Figure 4-29: *WOM*-plots for local and configural face manipulations.



To test all the assumed interrelations between features, a sequence of seven *T*-tests have do be calculated for the configural as well as for the local data. Table 4-18 gives an overview of all assumed interrelations of each processing model in terms of *WOM*. Additionally, the results of one-tailed *T*-tests are given to investigate whether these interrelations are in line with the respective model. All results which are in accord with the predicted relation are indicated by a special symbol. The structure of the table is the following:

- In the first column, the tested face pair is given, for instance 'E,EM' compares the *WOM* value of E with EM.
- The next two columns, called *P(C)* and *P(L)*, give the *P*-values of one-tailed *T*-tests for *configural* and *local* faces, respectively.
- Every model (M1, M2, M3, M4) lists the predicted differences of features in terms of *WOM*:
  "=": both features have an equal *WOM* score;
  "<": the first member of the pair has a lower *WOM* score than the second one;
  ">": the first member of the pair has a higher *WOM* score than the second one;
- For every model, there are two additional columns shown, one refers to *(C)onfigural*, the other refers to *(L)ocal* compatibility with the assumptions of the respective model. The following symbols are used:
  "++" : full confirmation of model assumption by empirical *WOM* results;
  "+" : ambivalent result (P-value is around the α–criterion of .05, i.e. .05 – .059);
  "–" : no confirmation of model assumption by empirical *WOM* results.

This schema lends strong support for Model M1 with the local data set, whereas Model M4 is favored by configural data (local data fit to Model 1 in *all* relations; configural data with a wider α–criterion of .059 in *all* relations, too). The overall accord of empirical data with model assumptions is indicated by bold typeface in blue in Table 4-17.

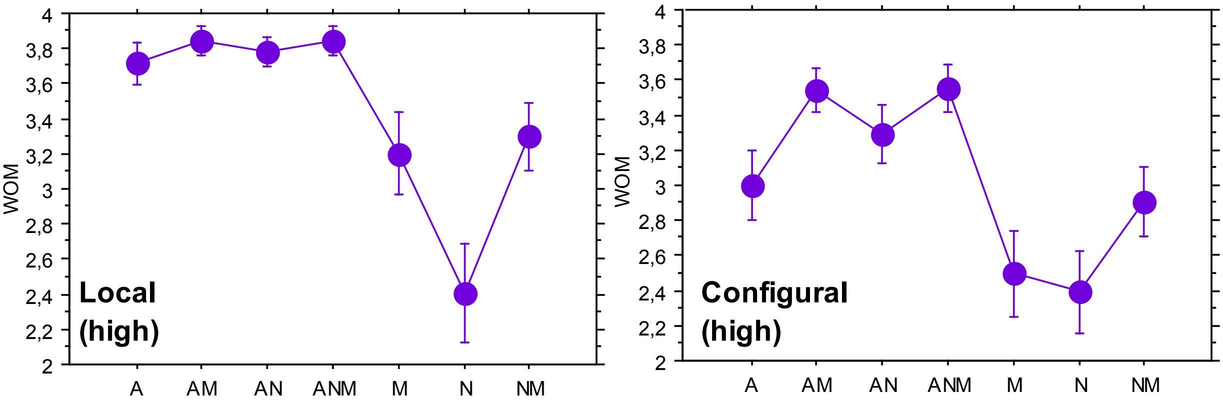
Table 4-17: Model testing. One-tailed *T*-tests (*df*=27) for low *and* high salience faces together. The first column shows the name of the analyzed pair, the second and third the *P*-value of the *T*-test for *(C)onfigural* and *(L)ocal* faces, respectively. The following columns show the model-predictions made by M1-M4 and whether the statistics referring the *P*-value fit to these model predictions. Accord with the model is indicated by a "++"-symbol; contrasting results by a "-". More information is given in the text above.

Pair	*P*(C)	*P*(L)	**M1**	C	**L**	M2	C	L	M3	C	L	**M4**	**C**	L
E,EM	.0004	.0014	=	++	++	<	++	++	<	++	++	<	++	++
E,EN	.0589	.2720	=	+	++	<	+	–	<	+	–	<	+	–
E,ENM	.0001	.0027	=	–	++	<	++	++	<	++	++	<	++	++
E,M	.0528	.0245	>	+	++	>	+	++	=	+	–	>	+	++
E,N	.0027	.0001	>	++	++	>	++	++	=	–	–	>	++	++
M,N	.0673	.0001	>	–	++	>	–	++	=	–	–	=	++	–
M,NM	.0131	.1093	=	–	++	<	++	–	<	++	–	<	++	–

In this calculation, high as well as low salience conditions were included. As was shown above, the low salience changes were very hard to recognize, especially for configural manipulations.

Therefore, to exclude floor-effects, the models were additionally tested with only *highly salient* faces in a second analysis. This will exclude artifacts of weak distinctiveness and extremely low recognition performance. The corresponding *WOM* plots for local and configural *highly* salient faces are both shown in Figure 4-30.

Figure 4-30: *WOM* data for local and configural *high* salience faces.



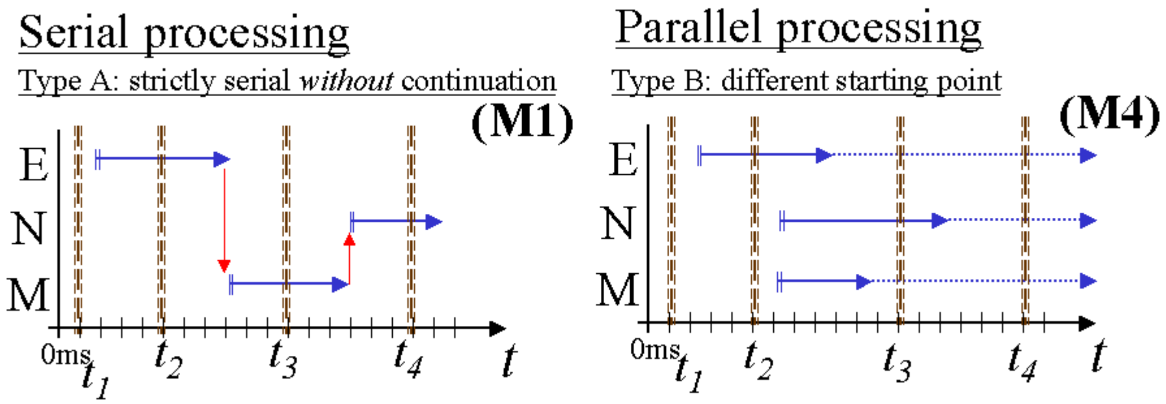
Using only highly salient faces, the model predictions from above were again calculated to test the model fits (see Table 4-18).

Table 4-18: Model testing. One-tailed *T*-tests (*df*=27) for only *highly* salient faces. The structure of this table follows that explained for Table 4-17 and in the text above.

Pair	*P*(C)	*P*(L)	**M1**	C	**L**	M2	C	L	M3	C	L	**M4**	**C**	L
E,EM	.0044	.1546	=	–	++	<	++	–	<	++	–	<	++	–
E,EN	.0541	.2052	=	+	++	<	+	–	<	+	–	<	+	–
E,ENM	.0027	.1978	=	–	++	<	++	–	<	++	–	<	++	–
E,M	.0235	.0443	>	++	++	>	++	++	=	–	–	>	++	++
E,N	.0141	.0001	>	++	++	>	++	++	=	–	–	>	++	++
M,N	.3268	.0020	>	–	++	>	–	++	=	++	–	=	++	–
M,NM	.0204	.2500	=	–	++	<	++	–	<	++	–	<	++	–

The *P*-values were marginally different from the full data, but the conclusions are approximately the same. Again, the local data favors Model M1 and the configural data favors Model M4. Figure 4-31 illustrates how local and configural changes are probably processed.

Figure 4-31: Most probable models for *local* (left side) and *configural* (right side) data.



A strictly serial model fits perfectly to the local data, whereas a parallel model with different starting points fits better with the processing of the configurally changed faces. The serial processing of features is also found with other paradigms and longer PTs of more than 300 ms (Posner & Cohen, 1984).

In this chapter, the terms *feature*, *face area*, or *face region* were used with an identical meaning. This is not self-evident, because the term *feature* describes an object-like entity, whereas the terms *region* and *area* have a more spatial connotation. It is questionable, whether the ‘features’ investigated in this experiment can be defined as real *objects* (E, N, M) or as the *location* of these features. The dissociation of objects and locations is well known in the field of short-term memory (STM) research, where recently much support was found for the contention that STM holds *objects* (or *features*) and *not* purely spatial areas (Lee & Chun, 2001). This was corroborated further by results from *IOR*-studies[50] (Tipper et al., 1991; Gibson & Egeth, 1994). Therefore, in this chapter the terms *location*, *area* or *region* will always be used with an objectural connotation.

Testing processing models with RT data

The above models were tested by *WOM* values, which are problematic in so far as this measure is new and therefore not well studied. To further test the results, an additional analysis of RT data was conducted. If it were the case that a strict serial processing underlies the recognition of local featured faces, or as Sergent (1984b) calls it, if an analytical process would occur, then the following RT pattern is predicted. The fastest RT to discriminate between faces could not be faster than latencies to discriminate between faces differing in the more salient feature. In terms of the above-described experiment, for example an EM-face could not be faster than an E-face. Contrary to this, for a parallel processing as assumed and demonstrated for configural faces, there has to be a difference between RT-data of EM- and E-faces. Within a parallel model, it is assumed that the more information is changed, the higher the probability to detect changes early. To test this prediction, two independent two-tailed *T*-tests with the difference of RT between E and EM for both manipulation classes were conducted. Both *T*-tests were contrasted to a hypothesized null-difference (see Table 4-19).

---

[50] IOR: *inhibition of return*. In studies of exogenous attentional orienting, response times for targets at previously cued locations are often longer than those for targets at previously uncued locations (Lupianez, Milliken, Solano, Weaver, & Tipper, 2001). Posner, Rafal, Choate, and Vaughan (1985) coined the term *inhibition of return* (IOR) to describe the effect.

Table 4-19: RT data for E- and EM-faces (RTs for correct trials; static outlier criterion of 300-5000 ms).

	RT(E)	RT(EM)	RT(E)-RT(EM)	*T*-value (*df*=27)	*P*-value
Configural	864.7 ms	771.7 ms	93.0 ms	2.87	.0079
Local	719.3 ms	703.0 ms	23.3 ms	.92	.3658, n.s.

As shown by the inferential statistics, there was no statistical difference between local E and local EM faces. This is accord with the assumption of a strict serial processing hypothesis. In contrast, there was a significant effect of features changed to a greater degree (EM) versus a single changed eyes region (E) for configural stimuli. This adds to the already reported evidence of a parallel processing assumption.

These *T*-tests were calculated for RT data with a static RT-outlier criterion. RTs had to lie inside an RT frame of 300-5000 ms (see Table 4-19). The analysis of RT data is tricky, because outliers can bias the results significantly (see section 4.2.2.2). Therefore, a second outlier criterion was defined, which was implemented in a dynamic way. For the second outlier criterion the RT range was defined by ±4 *SD*s around the arithmetic mean of the RTs of every single subject. Analogous to the data with the static criterion, again only the RT-difference between *configural* E and EM was significant (see Table 4-20).

Table 4-20: RT data for E- and EM-faces (RT for correct trials; dynamic outlier criterion of ±4 *SD*s).

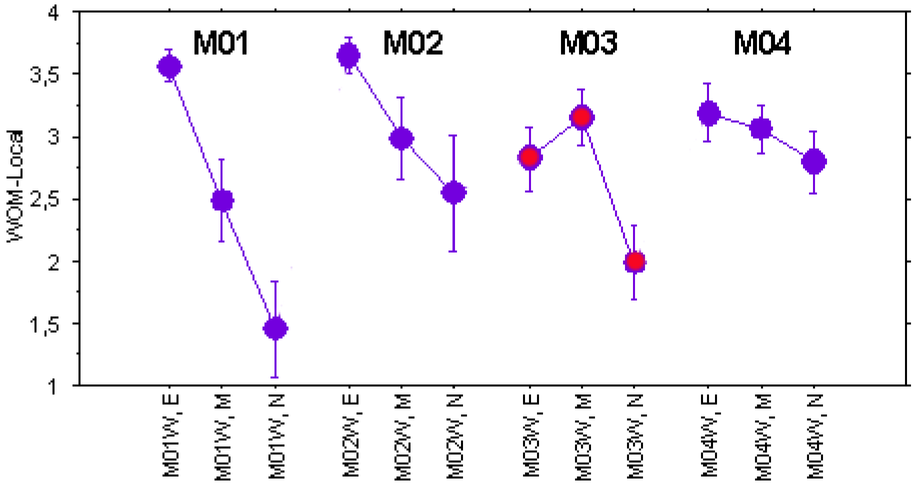
	RT(E)	RT(EM)	RT(E)-RT(EM)	*T*-test	*P*-value
Configural	864.7 ms	771.7 ms	101.5 ms	3.06	.0050
Local	726.3 ms	703.0 ms	16.3 ms	.74	.4689, n.s.

Again, the local data seem to be processed in a serial or analytical way. However, an analytical processing of local features faces was not found in a study by Sergent (1984b). The reason for this contradictory finding might be that Sergent used only one stimulus exchanging method and not a dissociate exchanging technique with either configural *or* local alterations. It is likely that a confoundation of the two methods might have caused her results. Sergent's (1984b) findings are more similar to the configural data set above. Therefore, the conclusion drawn by Sergent (1984b), namely that information from several features was combined interactively before a decision was made, is not valid for faces *per se*. Referring to the data above, it can only be applied to the configural face set, but *not* to the local one!

Empirical exceptions from the models discussed above

Concerning the sequence of feature processing, however, there are important constraints to the fixed order of E→M→N for local faces. As the local face-data further revealed, there were two important exceptions from the sequence found above of eyes first, then mouth and nose last. Experiment 1 used a between-subjects design in which four different basic faces were used in a between-subjects design. All faces except face M03 followed the sequence described above. For face M03, the analyses revealed an exceptional sequence. For this face, the mouth was recognized first and not the eyes, demonstrated by *WOM* data in Figure 4-32.

Figure 4-32: Local *WOM* data for the main features E, M, and N only. Face M01, M02, and M04 follow the trend discussed above of eyes first, then the mouth and the nose last. Contrary to this order, for M03 (red-blue points), the mouth was recognized before the eyes (indicated by a higher *WOM* value for the mouth).



It is obvious that all but face M03 followed this trend, with M01 revealing the clearest and M04 the most modest trend. Besides the exchange of the E and M position, the N position was stable over all faces. A closer look at the highly salient face material (see Figure 8-4 in the appendix) reveals that there was a construction error made for face M03. By accident, the highly salient mouth of face M03 was colored with a bright red lipstick, which stood in stark contrast to the relatively pale skin color. This ‘unnatural’ extreme distinctive cue is a plausible explanation for the reversed order found for this face. It seems that a very salient feature has the power to change the common feature processing sequence.

What might such a mechanism be useful for? It seems that locally changed features, which are not particularly salient, are processed in a fixed order of E→M→N. Nevertheless, if there are any unusually salient features, this sequence can be modified or re-programmed. The reason for this is to process most significant points before any others. Such a mechanism might be very useful to react to uncommon and new stimuli, which are suspected of being highly informative. This seems to be in accord with the results found by Yantis and Jonides (1990), who suggested that there can be top-down control of attentional priority, even to new objects forming rapid light onsets. Furthermore, Folk, Remington and Johnston (1992) demonstrated that participants were *not* able to ignore irrelevant information along the target dimension when this information was salient. Thus, even if non-targets are not helpful to improve the processing performance, their *salience* seems to influence the processing essentially—like the signal-red mouth of face M03.

In addition to this exception from the revealed sequence of locally processed features, one single participant also changed this order without working on the problematic face M03. Subject NG (subject nr. 20) showed a processing order of local features very similar to those participants that were presented to M03-faces with an extremely distinctive mouth. The post-experimental interview revealed that subject NG always focused the mouth region and mainly forced his processing to begin at the mouth region by volition. This is second evidence that the common feature processing sequence can be adapted for special reasons and is *not* hard-wired and cognitively impenetrable.

Concerning the investigations of sequentially processing of features, an important methodical point needs to be discussed. The testing of hypotheses on the subjects’ level has demonstrated that some effects might be covered within analyses of grouped data. Averaging over participants is indeed an important method to investigate general psychological phenomena. Nevertheless, averaging can also conceal interesting individual behavioral data (Loftus, 1996). Therefore, it is very useful to analyze on the individual level *beside* the grouped level, especially when processing hypotheses are tested. The failure to employ this additional method of analysis is also the weakness of Matthews' (1978) *Identikit* study. Matthews showed within a *same-different* RT-paradigm that changes to hair, eyes, or chins were detected equally fast and more rapidly than changes to eyebrows, nose, and mouth, the latter with increasing latencies. Therefore, these data are partly in accord with the results above. But as Sergent (1984b) correctly criticized, only a subject-based analysis could explore whether subjects scan the eyes, the chin, and the hair in parallel, or whether it has been a covered serial strategy.

The construction error of M03 has demonstrated that a very salient mouth can change the recognition process of a face. Other authors have already demonstrated that salience enhancement improves the recognizability of such features (Leder, 1996). It remains a question for future experiments whether it is possible to reverse the sequential processing of features with noses of increased saliency. By increasing the contrast and luminance, for example, the nose could become more important than the mouth. This is not a trivial question, because noses seem to be less salient *per se* than mouths or eyes (in frontal views), revealed for example by distinctiveness rating data in Pre-Study 1a. Therefore, artificially increasing the contrast or color of a nose might appear rather unnatural. Normally, noses[51] in frontally viewed faces have only a low cue quality. Additionally, noses play nearly no role for social communication, due to their relatively static nature. Therefore, the nose area seems not to be very important for a fast face recognition process. Thus, the increase of nose distinctiveness seems to be a challenging test, whether the sequence processing facial parts can be prioritized in favor of an 'unimportant' nose. An essential question would also be whether the character of the underlying process remains strictly serial under these circumstances.

Binding of configural and local processing

One other interesting question is how the two different processing characters (serial and parallel) are *binded* together (cf. von der Malsburg, 1981). How might the interfaces and the dependencies between the serial feature-by-feature and the parallel processing be structured? Binding is essential for building a whole and coherent facial image.

A typical two-stage model (e.g., Harner & Gaudiano, 1997; Treisman & Gelade, 1980) assumes that in a *pre-attentive stage*, parallel processing extracts local features. In a second *attention stage*, serial processing performs a more detailed investigation in areas of specific interest. From the empirical results of Experiment 1, such an account seems to be validated. The recognition data show that within the first 32 ms, a strictly serial processing of local features begins. The process starts with identifying the eyes. Subsequently, the mouth is recognized, and, lastly, changes to the nose are detected. Interestingly, this sequence of identification does not follow a simple top-to-bottom order! Moreover, the attentional control operates from the top (eyes) to the bottom (mouth), and then reaches the middle of the face (nose). The general processing order of features seems to reveal a pre-programmed top-down driven process with the flexibility to adapt to extremely salient points in the face. For particularly salient features, the order of the serial process can be changed. After about 63 ms and in unison with the serial recognition process, a configural or holistic identification begins. This second quality of recognition seems to be processed simultaneously.

Additionally, there were indications for a preceding separate processing stage. This stage was not explicitly addressed in the experimental design, but qualitatively explored in post-

---

[51] This can only be concluded for the frontal sight of faces as realized in these experiments. There exists other evidence from experiments with ¾-viewed faces for which the nose—due to its strong three-dimensional figure—played a more prominent role (e.g., Bruce et al., 1987).

experimental interviews. Therefore, it is rather speculative, but might merit further study in subsequent experiments. The participants reported that they sometimes had only a ‘feeling of seeing a face’, but not with any specific characteristics. It would be plausible to assume a very early stage of *coarse categorization*. Such a stage might operate on a very rudimentary level and is supposed to trigger off a successive finer processing. An important aim of this stage might be to localize the eyes for an orientation-normalization routine (cf. Figure 3-3 for an illustration in section 3.1.1). However, because neither recognition data nor RT data are available to validate this stage, this should only be mentioned here as a speculative assumption.

Influential factors for priority of processing

Another main result of this experiment is that salience *per se* is not the most important predictor for the detectability of changes in a face. The rationale behind the detectability is not a pure *quantitative* but rather a *qualitative* one. Therefore, local manipulations of the kind used in this experiment have a greater impact on the recognition of the face. The important question is whether this result is caused by the nature of processing as proposed above or whether it is only an artifact. It might be possible that the local-to-configural superiority is only due to a different size of manipulation for configural and local alterations. In Pre-Study 2, it was only examined whether *whole* faces with either local or configural manipulations differ from each other in terms of rated distinctiveness. Nevertheless, it was not tested whether the single areas in a *part*-based presentation differ from each other. Probably, this would be a more direct and adequate test for an equivalent operationalization of ‘commensurably’ configural to local alterations. For example, for short PTs it is conceivable that the cognitive apparatus might only perceive a small section of a face. Thus, not the salience of the full face but that of the section in question might be relevant for processing.

In order to investigate whether configural changes are really an equivalent manipulation or whether they rather correspond to local changes of low salience, a post-study (Post-Study 1) was conducted. In this study, not only full faces had to be evaluated in terms of distinctiveness, but also part-based versions.

4.3.5 Post-Study 1: Distinctiveness of part-based faces

Introduction.

The aim of this study was to further analyze the character of configural manipulation used in Experiment 1 (see p.76 et seqq.). Results from Experiment 1 showed a precedence of local information processing as opposed to configural processing. Distinctiveness is a valid predictor for recognition performance as well as for the speed of processing. To exclude that the results from Experiment 1 were due to the lower distinctiveness of configural information and are not based on specific sequences of processing, an additional distinctiveness study was conducted. It seems that in Pre-Study 2 these contingencies were already ruled out, because the distinctiveness ratings for configural and local faces were not found to differ. However, the conclusions from the results of Pre-Study 2 were drawn from ratings of *full* faces. The systematic manipulations of faces in respect to configural and local alterations have not yet been evaluated as *part-based* presented versions. Therefore, it remains unclear whether ratings of part-based presentations might not be a more valid instrument for testing distinctiveness of single features.

Thus, the material used for Pre-Study 2 was again rated, but in addition to the original study isolated features were also rated this time. The eyes region seems to be the only face area which might be evaluated in a full face presentation as well in a part-based one. A configurally changed nose or mouth is only presentable in combination with other juxtaposed features. Without them, the referential frame would be lost. This is not the case for configurally-changed eyes if the manipulation is realized by shifting the eyes inwards. The eyes region inherently contains its own reference system. In the following study, configural and local full faces had been compared with part-based versions. Therefore, only manipulations of the eyes region were tested.

Salience

The degree of salience of local features used for the artificially constructed faces of Experiment 1 have been evaluated in Pre-Study 1a. Therefore, distinctiveness ratings for the overall saliency of the newly constructed faces are assumed to validate these data. It is predicted that highly salient features are rated as highly distinctive and that those of low salience will be rated as being of low distinctiveness. This can be seen as a validation check of salience operationalization and of whether the combination of high/low-salience features within one face yield high/low distinctiveness for the whole face.

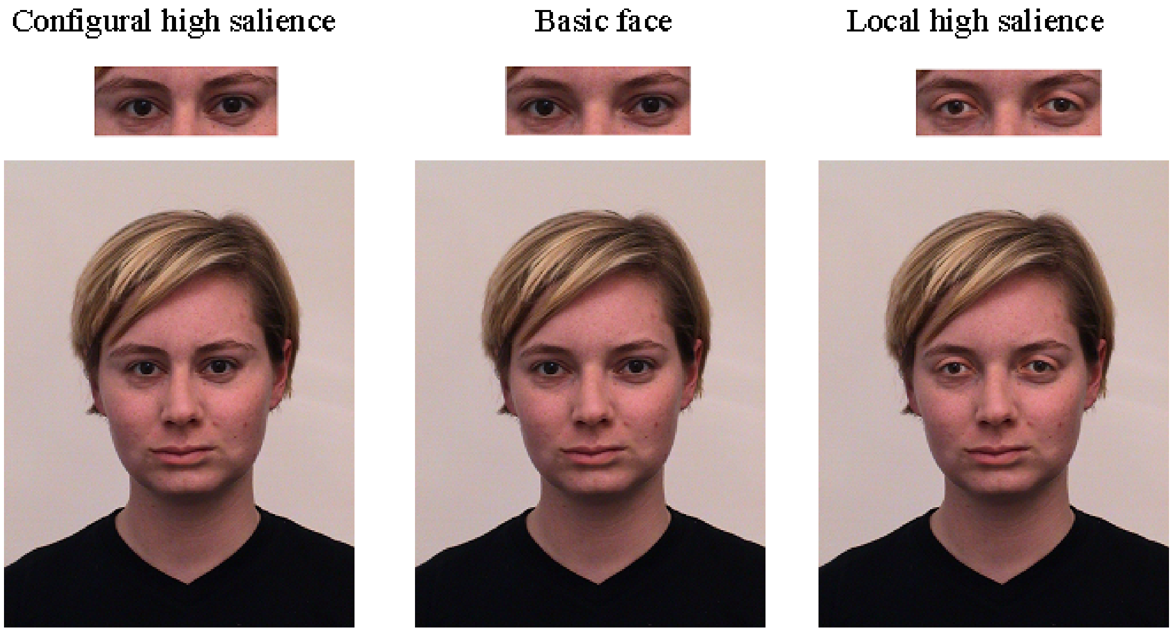
Face class

The distinctiveness ratings for configural and local faces are predicted not to differ. Analogous to the results of Pre-Study 2, it is assumed that there is no difference between both face classes. A dissociate distinctiveness score would be problematic for the analysis of recognition rates in Experiment 1. Highly salient faces are supposed to increase the recognition performance and shorten the RTs. Therefore, processing assumptions have to be based on face material with comparable salience scores.

Part-based salience

A part-based presentation of a face does not contain the whole information of a full face presentation. Most significantly, a fractional stimulus includes no global (and contextual) relations. The presentation of the eyes region, for example, does not give information about the configuration of the eyes in respect to the position of the nose or the mouth. Nevertheless, in the present studies the configural manipulation of the eyes was achieved with a technique that even a part-based presentation contains the (micro-) configural information. By shifting the eyes inward, the specific configuration is *not* lost. Theoretically, however, it seems more difficult to evaluate the distinctiveness of a fractional configural stimulus in contrast to the presentation of the whole stimulus due to the loss of *macro*-configural information (see Leder & Carbon, subm.-c). Figure 4-33 illustrates this reflection. Therefore, it is assumed that due to the different containment of global configural information, there is a discrepancy between the size of the presentation and the face class in terms of distinctiveness. Part-face structures are predicted to be rated as being of lower distinctiveness than full faces.

Figure 4-33: Presentation of full (lower row) and part-based (upper row) versions of a face. From left to right: Configural high salience manipulation, basic face, and local high salience manipulation. It is obvious that the configural changing needs the reference frame of the full face to be (recognized and) rated accurately.



Hypotheses.

There were three different hypotheses to be tested in terms of distinctiveness ratings.

- Hyp.1–*Salience*: The variation of (local) SALIENCE was operationalized by the data generated in Pre-Study 1a. Therefore, distinctiveness rating should replicate these results with lower scores for low-salience and higher scores for high-salience material.

- Hyp.2–*Face classes*: It is expected that there are no differences between configural and local faces in terms of distinctiveness. This is an important premise for testing face recognition and particularly the underlying face recognition *processes*.

- Hyp.3–*Part-based salience*: Configural changes presented in a part based modus (only eyes region shown) are more difficult to process. This might be due to a missing reference system. A reference system functions as a kind of holistic precondition that enables us to reveal larger configural structures. Therefore, lower distinctiveness ratings for configural faces presented in a part-based fashion are expected.

Method.

Participants. Sixteen students (12 women) of the Freie Universität Berlin, all of them undergraduates, participated in the experiment. They were given a course credit to fulfill course requirements. The mean age was 25.6 years (from 19 to 35 years). All participants had normal vision abilities or were corrected to normal vision.

Material. As material, all faces from Experiment 1 (see p.70) with manipulated eyes regions were used. These were 4 [basic faces] x 2 [CLASS: configurally vs. locally altered] x 2 [SALIENCE: low vs. high] = 16 faces. Additionally, there were two SIZE variations. For the *eyes-condition*, the eyes region was cut out. The *full-condition* showed full faces as described in Pre-Study 2.

Procedure. The procedure was similar to the distinctiveness-rating task in Pre-Study 2 (see above), but now all four basic faces were used within subjects. The total number of trials was 2 [SIZE: full faces vs. eyes-region only] x 16 [faces] = 32. The SIZE factor was realized by a blockwise presentation, balanced over subjects. The whole procedure lasted about 8 minutes.

Results & Discussion.

The mean distinctiveness rating data are shown in Table 4-21.

Table 4-21: Mean distinctiveness rating for all experimental conditions. In square brackets the *SD*s are shown.

CLASS	SALIENCE	SIZE=eyes	SIZE=full
configural	high	3.984 [1.17]	4.750 [.97]
local	high	4.766 [1.31]	4.828 [1.04]
configural	low	3.234 [.93]	3.750 [1.07]
local	low	3.219 [1.12]	3.828 [1.00]

In order to test the hypotheses, the distinctiveness data of each participant were submitted to a three-way repeated measurement ANOVA with CLASS (local vs. configural), SALIENCE (low vs. high) x SIZE (full vs. eyes) as within subjects factors (see Table 8-28 in the appendix for details).

Hyp.1–Salience

The analysis revealed a significant factor SALIENCE ($F_{1,15}$=63.50; P<.0001) with highly-salient features rated higher than low-salience features in terms of distinctiveness (Ms=4.58 vs. 3.51). This is a direct validation of the salience classification from the data of Pre-Study 1a, as assumed by hypothesis Hyp.1.

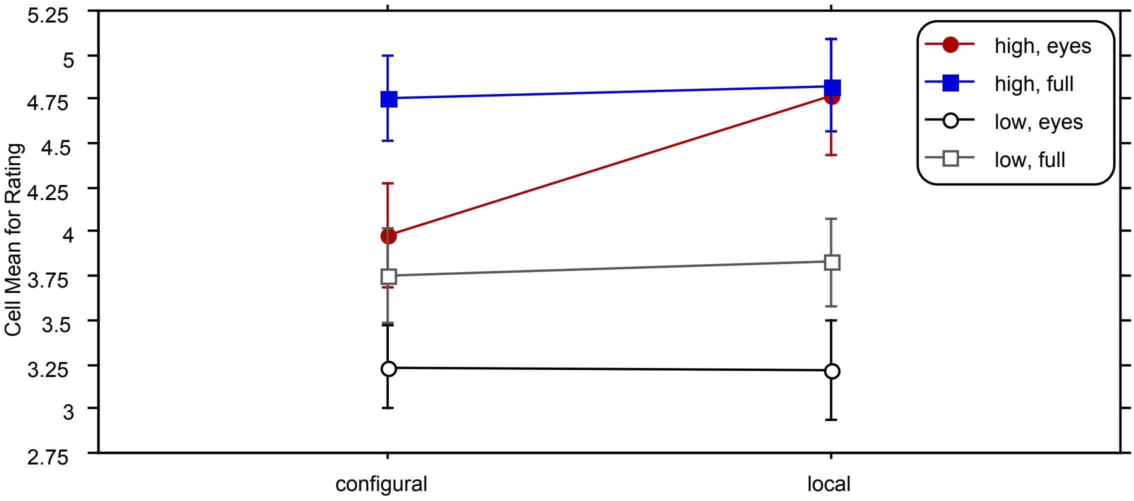
Hyp.2–Face class

The factor CLASS was not significant, but there was a strong trend of a higher rating for local versions than for configural ones (Ms=4.16 vs. 3.93, $F_{1,15}$=3.92; P=.0670). This result can not be interpreted reliably, because the power of the test, assuming a large effect of f^2=.35, was only .60 (λ=5.6; $F_{1,15}$=4.54) (Erdfelder, Faul, & Buchner, 1996), so the risk of wrongly accepting the null hypothesis lies at 40%.

Hyp.3–Part-based salience

The third factor SIZE also was significant ($F_{1,15}$=5.19; P=.0378), with the full face ratings higher than the eyes in isolation (Ms=4.29 vs. 3.80). Additionally, and most interestingly, there was a three-way interaction of CLASS * SALIENCE * SIZE ($F_{1,15}$=6.15; P=.0255). In Figure 4-34 it is shown that the trend of CLASS was not a *general* trend of advantage of local over configural changes.

Figure 4-34: Interaction of Class * Salience * Size in terms of distinctiveness rating.



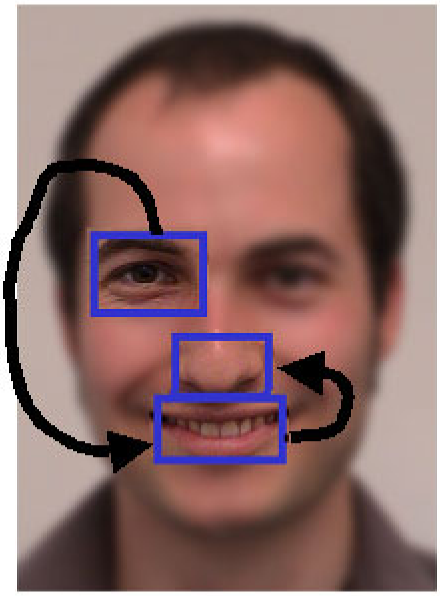
Moreover, configural compared with local changes were less salient in only one condition. The distinctiveness dropped significantly only when the stimulus was a highly salient configural one presented in *isolation* (condition EYES). Under this condition, there was a significant difference of the distinctiveness ratings (T_{15}=3.96; P=.0013) with M=3.98 (for configural highly salient isolated stimuli) and M=4.77 (for local highly salient isolated stimuli). However, why is this three-way interaction so important?

There was a general decrease of distinctiveness ratings when the rated objects were presented only as isolated parts. This effect was nearly the same for configural and local versions with low *a priori* salience, supporting Hyp.2. Parallel ratings are a good indicator for a harmonious and equivalent operationalization of configural and local versions. In comparable experimental studies (e.g., Rhodes, 1995), the degree of salience is normally *not* controlled *across* different manipulation classes, so it is difficult to conclude which effects really caused the results. Conservatively, it could be assumed that it is only a simple pre-experimental saliency effect. This is not the case for the present experiment. The empirical results of Pre-Study 1a, Pre-Study 2, and Post-Study 1 demonstrated that manipulations were very well operationalized for both face classes, but only when *full* faces had to be rated. This was no longer the case when highly salient facial structures were presented in a *part-based* version. Thus the drop of distinctiveness rating for these highly salient facial features partly support Hyp.3.

The dissociation of part-based vs. full face presentation was investigated further. Two additional repeated measurement ANOVAs with CLASS and SALIENCE as within factors were calculated for full-SIZE (see data in Table 8-29 in the appendix) as well as for eyes-SIZE (see data in Table 8-30 in the appendix). As in Pre-Study 2, factor CLASS was not significant for full faces ($F_{1,15}$=.17; P=.6860, n.s.). However, when presented in a part-based version (eyes region only) there was a significant factor ($F_{1,15}$=9.66; P=.0072). When the eyes-region only was presented, there was a difference between configural and local stimuli! Thus, the finding of Pre-Study 2 was replicated by the full-SIZE faces, but was refined by a significant CLASS factor for part-based stimuli.

To explain this dissociate effect in an analogy, the effect might be as if seeing idiosyncratic objects through a narrow window (for an illustration of this analogy see Figure 4-35). If only parts of a larger structure are available through such a window, highly salient local parts are still highly distinctive, because they are *inherently* salient with no need for contextual information to validate them. Local features, after all, can easily be recognized as 'wholes' themselves, even in isolation (cf. Meltzoff & Moore, 1977).

Figure 4-35: Illustration of face processing based on a focal window metaphor.



Therefore, concerning local features it does not matter whether they are seen in the context of the super-structure or in isolation, because they have a special quality of their own. Contrary to local features, configural substructures would have a different character of salience, because this kind of salience is seriously affected by the context and is therefore *not* an inherent quality of the kind of alteration. Surely, this could be the case for certain local changes as well. However, local configurations always change their contextual micro-configuration implicitly. For example, by exchanging a nose not only local features are affected, but also configural aspects due to the relations to other features (Leder & Carbon, in prep.).

This finding seems to be an important explanation of the general results of Experiment 1, especially the low detectability of configural changes within the first 30-50 ms. If it were assumed that the very early processing of faces recruits much from local scanning strategies which are not able to construct a whole facial picture within such a short time window, then this strategy is like a moving attentional window. This window will be placed over the most attentive areas (hair, eyes, nose, and mouth) due to their salience. However, within this narrow window configural alterations have a big disadvantage compared to local versions in spite of their psychological distinctiveness. Therefore, the attentional window will on the one hand be moved very quickly to the next area without deeply processing this area. On the other hand, the sensitivity for configural changes within such a frame might not be enough to discriminate the salience in a fine manner.

This analogy extends the results of Leder and Bruce (2000b), who supposed that configural information is *locally* processed and not as proposed here, in a more global manner. Leder and Bruce (2000b) investigated different constraints and a different task (*face recognition task*). Therefore, it would be worthy of further analysis under which conditions more *local* and under which conditions more *configural* features are processed.

Summarizing these results, the salience was operationalized for configural and local faces in a comparable way. When presented in full-size, both face classes did *not* differ in terms of distinctiveness (with the given low statistical power). This was not the case for part-based versions, in which configural alterations were rated as being less distinctive. It is important to consider these results in respect to interpreting the recognition data of Experiment 1. On the one hand, only *full* faces were used as stimuli in Experiment 1. In Pre-Study 2, *full* faces had to be evaluated in terms of their distinctiveness. It was found that there was no difference between configural and local faces in respect to distinctiveness. Therefore, it could be concluded that differences between the recognition of configurally and locally changed faces are exclusively due to the underlying nature of these processes. On the other hand, briefly presented faces might result in a more part-based processing. Thus, distinctiveness ratings of part-versions might be the more valid instrument to calculate the pre-experimental salience of a stimulus. However, Post-Study 1 has revealed that high-salience configural changes were hardly detected when presented in a part-based quality. Thus, there indeed seems to be a pre-experimental difference between configural and local faces in terms of salience.

4.3.6 Re-analysis of the Pre-Studies: What is distinctiveness based on?

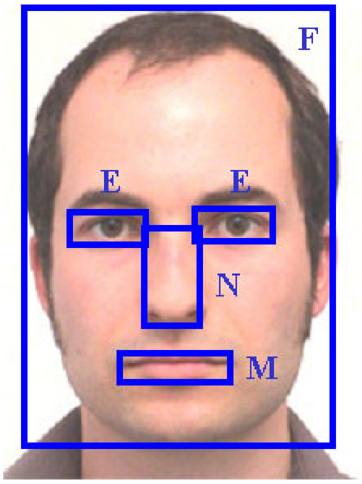
In Experiment 1, it was demonstrated that distinctiveness played an eminent role in the recognition process. Distinctiveness is known as a powerful determinant for memory performances in general and face recognition in particular (e.g., Bruce et al., 1993; Bruce, Burton, & Dench, 1994; Sommer, Heinz, Leuthold, Matt, & Schweinberger, 1995). Furthermore, distinctiveness, untypicality, idiosyncratic features, uncommon spatial relations, recognition performance, and processing speed are intimately related. Nevertheless, it remains unclear what actually determines distinctiveness psychologically. According to *Bloch's Law*, less distinctive features are also weaker physically (Loftus & Ruthruff, 1994). The basis of Bloch's law is that memory for complex visual stimuli has been discovered to improve with greater stimulus duration (Loftus & Kallmann, 1979; Shaffer & Shiffrin, 1972) as well as with greater stimulus intensity (Loftus, 1985). Loftus and Ruthruff (1994) have shown that with increasing intensity, a smaller stimulus duration is required for performance to exceed chance level. Contrary to this physical hypothesis, other researchers claim that most of the dimensions underlying similarity-of-face judgments and therefore distinctiveness are not *physical* factors (size of the nose, color of the skin, etc.). They believe that distinctiveness is rather based on global and inferential features. A prominent candidate for such a higher cognitive representation seems to be attractiveness (Hirschberg, Jones, & Haggerty, 1978). Therefore, attractiveness is also taken into account as a possible predictor of distinctiveness.

Featural base of distinctiveness

The following re-analysis of data is aimed at investigating what factors primarily determine distinctiveness. According to the physical distinctiveness hypothesis, there should be a strong interrelation between physical factors like size, color, shading, or contrast and the rated distinctiveness. A similar but much more elaborated approach was presented by Bruce, Burton, and Dench (1994). They took a large number of measurements from a set of 89 male and 86 female faces. For every face, all relevant featural sizes, distances, and angles were measured from a full-face view (e.g., nose length, mouth width). In addition, more complicated distances, ratios, and angles measured from a full-face and profile photograph were used (e.g., angle of nose bridge). The goal was to explore the inter-relationship between distinctiveness and these measures.

For the following analyses, the sole focus lay on the frontal view of faces, because all studies and experiments described in the present work used frontal images only. Therefore, the size of all relevant inner parts of the (frontal) faces was measured (eyes, nose, mouth). Figure 4-36 illustrates the measuring method of the eyes (E), nose (N), mouth (M), and the full face (F). The size measures of the eyes, mouth, and nose was also standardized to the total size of the square measure of the full face.

Figure 4-36: Illustration of the measuring method of the size of the features. To simplify the measuring method, the square limited by the most peripheral contours of these features (E, N, M) was taken as area measure. This measure was standardized to the total size of the square measure of the full face (F).



In order to ascertain how strongly the different features relate to distinctiveness, a correlation analysis with the (relative) size of features as independent and distinctiveness as dependent variable was conducted. The size of the eyes had a measurable influence on the distinctiveness of eyes (R=.572), which was not the case for the mouth nor for the nose. The correlations are presented in Table 4-22.

Table 4-22: Possible physical and aesthetic predictors of distinctiveness of the eyes, the nose, and the mouth (taken from Pre-Study 1a), as indicated by intercorrelations with the respective feature's distinctiveness.

Predictors	*Eyes*	*Nose*	*Mouth*
Relative size	.572	-.041	.197
Distance	-.073	.388	-.261
Luminance	-.039	-.052	.041
Relative luminance	-.349	-.203	.118
Attractiveness rating (full face)	.399	.131	.308

Configural factors

Another potential physical measure predicting the distinctiveness of features are configural aspects of a face. The inter-oculary distance of the eyes (for the nose feature), the nose-eyes distance (for the nose feature) and the nose-mouth distance (for the mouth feature) were analyzed in respect to predicting distinctiveness. This correlation analysis also failed to find a strong relationship with distinctiveness ($R_{\max}$=.388; for the nose).

Luminance

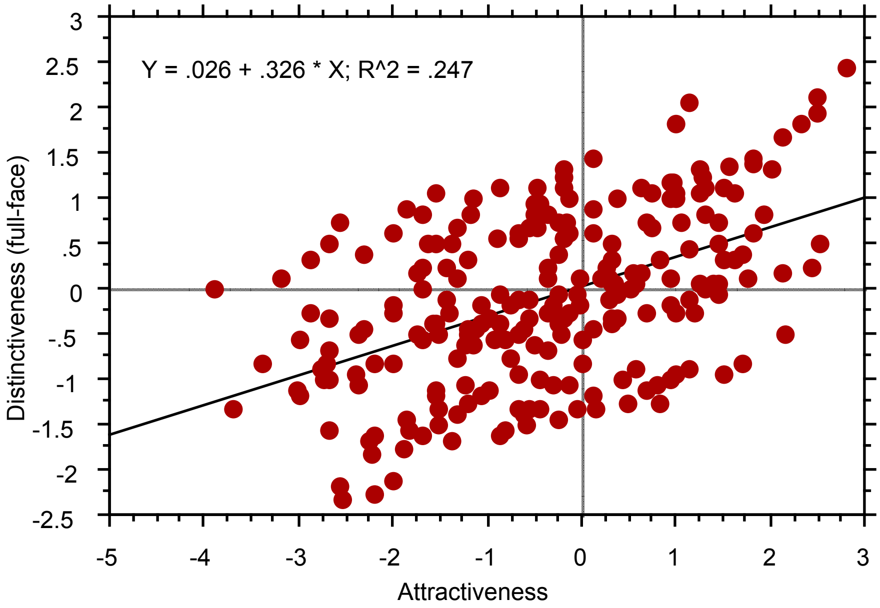
Another candidate as a possible predictor of distinctiveness was the luminance of the features. The luminance was used as a weighting of the color-components in relation to their physiological impact as the measure. Following Altmann (1999), the luminance was weighted according to the compatibility-scheme of color and black/white TVs. Thus, the resulting luminance is weighted with 30% of red, 59% of green, and 11% of blue. Interestingly, darker features (especially dark eyes) had the tendency of being more distinct than lighter ones (shown by a negative correlation). Again, there were no strong relationships ($R_{\max}$=-.052), even when the luminance was standardized to the overall luminance of the skin. The standardized luminance of features, which functions like a contrast indicator, correlated only modestly with distinctiveness (R_{max}=-.349; for the eyes). In Table 4-22, all these correlations are listed in one single survey.

Attractiveness

The attractiveness of the full face also revealed only a weak relation with distinctiveness of the single features. A correlation analysis indicated a modest relation of R_{max}=.399. Therefore, the concept of facial parts' distinctiveness is also relatively difficult to predict from attractiveness.

However, the supposition of Hirschberg et al. (1978) that the *overall* distinctiveness (of the full face) is more dependent on global and inferential qualities like attractiveness was partly confirmed. A regression analysis between attractiveness and distinctiveness of *full* faces revealed a stronger correlation between these two factors with R=.497 (F_{230}=75.00, P<.0001; data from Pre-Study 1a). This correlation was again not very high, but much higher than nearly all possible predictors tested above. Figure 4-37 shows the regression plot for attractiveness as independent and distinctiveness of full faces as dependent variable.

Figure 4-37: Linear regression analysis with attractiveness as independent and distinctiveness as dependent measure. Both measures are ratings for *full* faces from Pre-Study 1a.



Reiterating the results found above, neither physical measures nor higher cognitive variables were able to predict either the distinctiveness of features or the distinctiveness of full faces with more than 33% of explained variance (see Table 4-22).

4.3.7 Experiment 2: Elongation of PTs for unfamiliar faces

Introduction

Experiment 1 gave much insight into the processing of local and configural faces within the first 100 ms (exactly 32–94 ms) of presentation. Two important questions remain.

PT elongation

First, is it possible to elongate the experimental design of Experiment 1 in terms of presentation times (PTs)? The recognition rates in Experiment 1 did not reach the ceiling of 100%. Therefore, it would be important to investigate how the recognition continues to develop for longer PTs. In the following experiment, the PTs are set to 200 ms and 400 ms, respectively.

Recognition of the nose

Second, the recognition of altered nose areas was very poor. In the proposed face processing models FACEREC-1, M1 and M4, a late processing of the nose is assumed. It is questionable whether the late and inaccurate processing of nose-features is really caused by a low processing-priority of the nose or whether nose-alterations are generally hard to recognize, regardless of time resources. Therefore, in Experiment 2 the same procedure as in Experiment 1 was used, but with longer PTs. With either 200 ms or 400 ms, faces were shown much longer than in Experiment 1. Within such a time-period, nose alterations should be recognized much more easily than within short presentation times, presupposing that the nose area is processed relatively late.

Nevertheless, it is not warranted to interpret the data of Experiment 2 as a test-elongation, because firstly, visual processes with PTs over 200 ms are not pre-saccadic anymore (Julesz, 1981a; Kosslyn, 1994). Secondly, the data were not taken from the *same* participants as in Experiment 1. Thus, both data series cannot be tested in one overall within-subjects design. Therefore, this experiment investigates whether noses are ‘features that are hard to recognizable’ *per se*.

Hypotheses.

Two hypotheses were tested concerning the elongation of PTs in respect to the original PT setting in Experiment 1.

- Hyp.1–*Validation of PT elongation*: The accuracy of recognition for both PTs follows the trend of the first seven PTs of Experiment 1. Generally, it is assumed that the longer the PT, the higher the recognition rate.

- Hyp.2–*Recognition of the nose*: It is assumed that the low recognition of faces with altered noses (N) in Experiment 1 was not an artifact of a generically low performance in the recognition of noses. Moreover, it is caused by a late beginning of the processing of this featural area. Therefore, the correct percentage of the N-condition for both PTs is expected to be above the chance criterion of 50%.

Method.

Participants. 28 students (20 women) of the Freie Universität Berlin, all of them undergraduates, participated in the experiment. They were given a course credit to fulfill course requirements. The mean age was 24.6 years (from 19 to 38 years). All participants had normal vision abilities or were corrected to normal vision.

Material. The material was the same as in Experiment 1.

Procedure. The procedure was very similar to that of Experiment 1, but in Experiment 2 the PTs were either 200 ms or 400 ms, respectively. Thus, the test-block of the experiment contained only 2/7 (112) of the trials of Experiment 1. The whole procedure lasted about 25 minutes including instructions, training, and a post-experimental interview.

Results & Discussion.

To test both hypotheses, the recognition rates of manipulated faces were analyzed. In Table 4-23 the percentage correct data for all conditions of Experiment 2 are given.

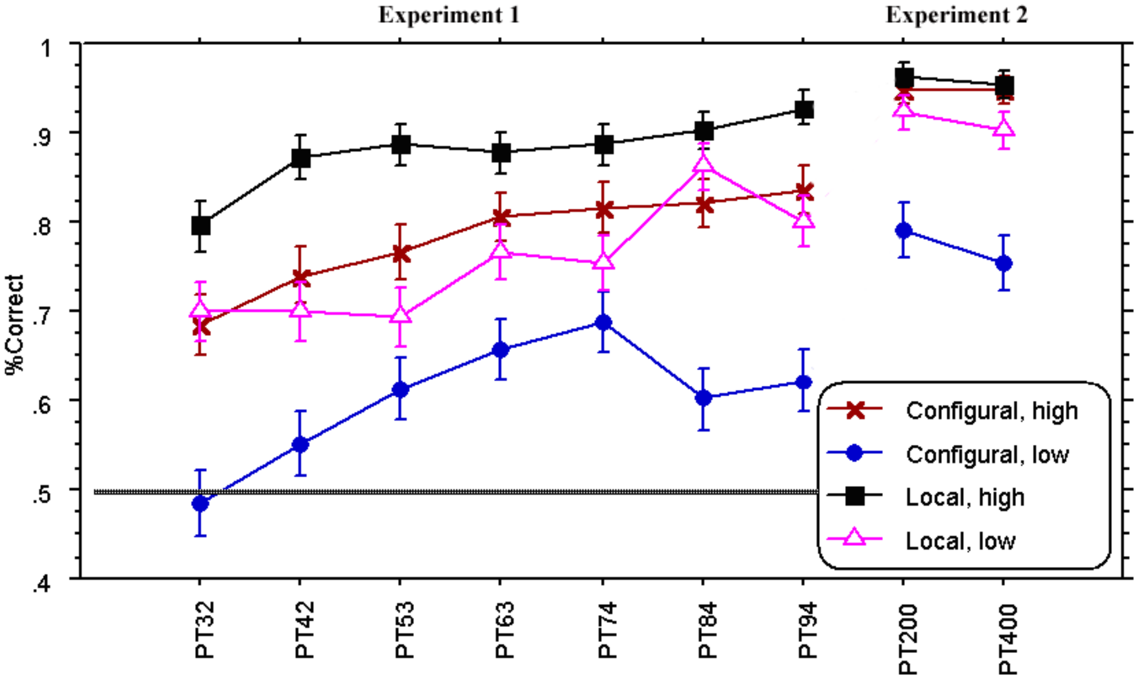
Table 4-23: Recognition rates for all conditions in Experiment 2. Additionally, in column '*PT94*' accuracy data for PT=94 from Experiment 1 are given (data in italics) to be able to compare the recognition data for the longest PT of Experiment 1 with data of Experiment 2.

CLASS	SALIENCE	FEATURE	*PT94*	PT200	PT400
configural	high	E	*0.714*	0.893	0.964
		EM	*0.964*	1.000	1.000
		EN	*0.964*	1.000	1.000
		ENM	*1.000*	1.000	1.000
		M	*0.714*	1.000	0.964
		N	*0.571*	0.786	0.750
		NM	*0.929*	0.964	0.964
	low	E	*0.571*	0.750	0.821
		EM	*0.786*	0.929	0.964
		EN	*0.643*	0.893	0.929
		ENM	*0.786*	0.964	0.964
		M	*0.607*	0.679	0.679
		N	*0.393*	0.536	0.393
		NM	*0.571*	0.786	0.536
local	high	E	*1.000*	1.000	1.000
		EM	*1.000*	0.964	1.000
		EN	*1.000*	1.000	1.000
		ENM	*1.000*	0.964	1.000
		M	*0.821*	0.929	0.929
		N	*0.786*	0.893	0.821
		NM	*0.893*	1.000	0.929
	low	E	*0.929*	1.000	1.000
		EM	*0.964*	1.000	1.000
		EN	*0.857*	1.000	0.964
		ENM	*0.964*	1.000	1.000
		M	*0.679*	0.964	0.964
		N	*0.464*	0.536	0.464
		NM	*0.750*	0.964	0.929

Hyp.1–Validation of PT-elongation

In Figure 4-38 the data of the recognition performances above are illustrated in combination with the referring data of Experiment 1. There is a clear trend for higher recognition at longer PTs. All percentage correct rates for the longer PTs from Experiment 2 lie above the performance of those of the longest PT (94 ms) from Experiment 1.

Figure 4-38: Accuracy data for CLASS (configural and local) * SALIENCE (low vs. high) faces with PTs from 32 ms up to 400 ms. The data stem from Experiment 1 (32 ms–94 ms) and from Experiment 2 (200 ms–400 ms), respectively.



Nevertheless, it is highly speculative to interpret these data as a pure test-elongation of Experiment 1, because the data of each experiment are based on within-subjects designs, but each with different subjects. Interestingly, all but the configural low-distinctiveness data reach a performance level of nearly 100%, which is very high for PTs of fewer than half a second. However, the low-distinctiveness configural faces were also recognized above the chance level of 50% (two-tailed T-test with T_{27}=9.50, P<.0001).

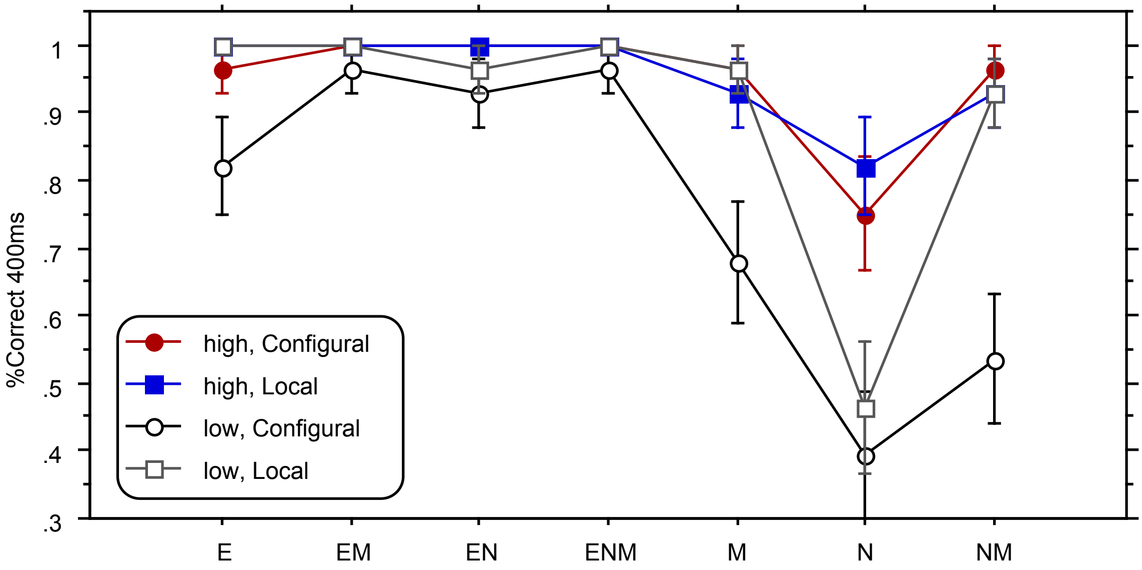
A PT of 200 ms seems to be advantageous in terms of recognition performance over those used in Experiment 1, but could additional 200 ms (i.e., PT=400 ms) be beneficial for further improvements of the recognition rate? To test this, a four-way repeated measurement ANOVA with CLASS (configural vs. local), SALIENCE (low vs. high), FEATURE (E,N,M,EN,EM,ENM), and PT (200 ms vs. 400 ms) as within factors (see Table 8-31 in the appendix) was conducted. All main factors except PT ($F_{1,27}$=1.75; P=.1966, n.s.) were found to be significant. This indicates a kind of saturation of recognition, because the performance level is still impressively high at 200 ms. Even an elongation of PT to 400 ms could not further increase recognition substantially.

Nevertheless, the continuation of the recognition trend from Experiment 1 supports Hyp.1 with the conceptual improvement that due to the asymptotic character, 200 ms are found to be sufficient to recognize changes in faces very efficiently.

Hyp.2–Recognition of the nose

Crucial for this experiment was the analysis of the nose condition, because changes to the nose were recognized only at a low level in Experiment 1. The aim was to investigate whether the detection of a face with an altered nose lies above the sensory threshold or whether the low recognition rates in Experiment 1 were only due to the short PTs. In order to test this, only the recognition rates with PT=400 ms were analyzed. A presentation of 400 ms seems to best resemble face recognition in everyday encounters. Figure 4-39 shows all feature combinations for both manipulation classes and the different degrees of salience, based on the data of Table 4-23.

Figure 4-39: Recognition rates for PT=400 ms. Each feature combination is shown under all CLASS * SALIENCE conditions.



Most obvious in this illustration is the weak recognition rate for the nose (N). This is also indicated by a significant factor FEATURE ($F_{6,162}$=31.78; P<.0001) in the above ANOVA. Most remarkable is the drop of recognition performance for the low-salience conditions. The recognition of the configural as well as the local nose manipulation—if changed in a low-salience direction—was not statistically different from chance (two-tailed T-tests: $T\text{s}_{27}$<1.14, Ps>.2642, n.s.). This in part falsifies Hyp.2.

Actually, this conclusion is not unproblematic for the data analyses in Experiment 1, because it seems that noses altered in the direction of low distinctiveness are *beneath* the sensory threshold and therefore could bias the analyses of processing. Nevertheless, the models in Experiment 1 were tested with either *all*-distinctiveness faces as well as with *highly* distinctive featural manipulations, but *not* with low distinctiveness alone. However, for the highly salient conditions, changes of the nose area were detected (two-tailed T-tests: $T\text{s}_{27}$>3.00, Ps<.0057, n.s.). Thus, the processing-conclusions drawn for Experiment 1 are still valid.

For further experiments with similar stimulus material, the weak recognition performance of low salience noses must be taken into account. It seems more sensible to use only the high-salience faces and none of the low salience faces to prevent this problem. A less useful idea would be to generate faces that are even more distinctive, because the plausibility would decrease dramatically with such uncommon faces (see plausibility ratings for highly salient configural faces in Pre-Study 2).

4.3.8 Summary of studies and experiments with unfamiliar faces

In the previous sections, the early processing of faces was investigated. There were two major aims. On the one hand, the nature of pre-saccadic feature processing was analyzed. Four different processing models were tested. On the other hand, the impact of different information qualities in faces on the nature of the feature processing was investigated further.

In order to test the nature of processing, systematically manipulated faces had to be used within a *change detection* paradigm. Thus, artificial faces were constructed (Pre-Study 2) from components of systematically selected natural faces (Pre-Study 1a). There were three main factors which were controlled for these faces. First, the featural factor with systematically altering the eyes-, nose-, and mouth-areas (E,N,M). Second, the kind of alteration was varied (CLASS). On the one hand, local features were replaced by alternative local features from other faces ('local' class). On the other hand, there was a 'configural' class, for which the inter-relations between the single features were changed. Third, the salience of these features was varied as low- or high-salience.

These artificial faces (*DADA-M*-faces) were presented in Experiment 1 in a sequential matching paradigm with limited presentation times (PT) from 32 ms–94 ms, followed by a visual mask. The participants had to decide as fast and as accurately as possible whether two following faces were the same or different ones.

Salience emerged as a very important factor in the quality of recognition. High-salience featural changes were detected faster and more accurately than low-salience ones. Moreover, the eyes-region was found to be a highly distinctive feature. Changes to the eyes were always recognized very fast and with a high rate. In contrast, the nose area was hardly recognized, regardless of the way it had been altered.

Interestingly, the class of feature changing influenced the processing of features. The features of locally changed faces seem to be processed in a strictly serial way, whereas configural features were processed in parallel. Typically, the strict serial processing of local features started with E, then processed M and lastly N (E→M→N). This sequence seems to be rather counter-intuitive, because a simple top-to-bottom strategy (E→N→M), for example, would be much easier to implement. There were only deviations from this processing sequence of features when the face contained an extraordinarily salient feature, like an obtrusive-looking mouth. Then the sequence was found to prioritize this most salient point in the following way: M→E→N.

In Experiment 2, different PTs were used (200 ms and 400 ms, respectively). This PT elongation was applied to investigate whether the poor recognition rates for changed noses were due to generically poorly detectable nose changes or to late processing of the nose area. High-salience noses were found to be recognizable within a PT of 400 ms, which supports the hypothesis of late processing of noses.

Both experiments showed that all features of a face were processed not *at once* or in terms of a holistic processing. Moreover, the *quality* of the features and their *salience* was essential for their processing. The processing models discussed for these experiments were tested with an integrative measure called *WOM* as well as with conventional recognition rates and reaction times. Nevertheless, further work has to be done to investigate the different processing types and their causes in more detail and with alternative methods.

4.4 Experiments with familiar Thatcher faces

This experimental series studied the involvement of several recognizing strategies in the early processing of faces. The time span of interest for this processing were the first 200 ms of the presentation of a face. Three main aspects have been investigated. First, the importance of local feature analysis for early processing. Second, the specific role of global cues. Third, the time at which configural processing comes into play to bind several features.

Participants had to recognize famous faces within a short presentation time (PT). In order to test local as well as configural processing, so-called *Thatcher face*s seem to be an appropriate stimulus material. In Thatcher faces, the mouth and eyes region is turned upside-down. When inverting the whole face, the strong alteration of the face is usually not detected. However, due to this specific construction of Thatcher faces, the local regions of the eyes and the mouth are then again oriented upright. Therefore, it is possible to test, whether this locally correct orientation is advantageous for the recognition process. Consequently, Thatcher faces were used as the main material to test different processing hypotheses.

The structure of this experimental series is tripartite.

Experiment 3a and Experiment 3b

In the first part, two experiments investigated the special role of local feature analysis in early processing. Additionally, it was examined whether the amount of local feature analysis changed over the PT-span from 26 ms to 200 ms. Furthermore, the detection of facial incoherence in faces was studied.

Experiment 4

The next experiment tested whether the results of Experiment 3a are also valid for an alternative visual mask for limiting PTs. Additionally, further hypotheses concerning global processing were examined.

Experiment 5

Experiment 5 used an alternative interrogative form in the test. This was done in order to enable generalization of the results to a more realistic face recognition task. Again, local as well as global processing hypotheses were tested.

4.4.1 Experiment 3a: Inverted Thatcher faces

Introduction.

The aim of Experiment 3a was to study early processes of face detection and face recognition with the focus on local identification processes. In the previous chapters it was demonstrated that local recognition processes start very early. Nevertheless, it was not investigated whether it is *advantageous* for the recognition of a whole face to identify local features very early. In order to test this, so-called *Thatcher faces* were used as stimulus material. See section 3.1.3 for details about Thatcher faces. Thatcher faces are constructed by turning the area of the eyes and the mouth upside-down. After this manipulation, the face looks very grotesque and does not resemble the original person very much. However, by turning the whole picture upside-down, the faces lose their grotesqueness and can again be identified as the original pictures (for an illustration of the Thatcher-illusion see Figure 3-6 in section 3.1.3). There have been some explanations for this strong perceptual effect. These are described in section 3.1.3.

In the following paragraph, inverted Thatcher faces are used as a tool to study specific processes of face recognition. One explicit advantage of inverted Thatcher faces over other face classes is that Thatcher faces are rather indistinguishable to (inverted) original faces, although they are strongly altered. The following experiment will be concerned with the following aspects of face processing.

Thatcher detection

Thatcher faces are physically very different from normal faces. By inverting Thatcher faces, this obvious difference is nearly lost perceptually. However, it must be investigated whether there are differences between the recognition of Thatcher faces and original faces. It is predicted that Thatcher faces are sometimes detectable as being thatcherised, i.e. as being odd.

PT effect

In these experiments, not only early processing of faces was tested. Moreover, the *development* and *change* of processing were scrutinized. Therefore, a variation of PTs was used. There is a general recognition advantage for longer PTs (see Laughery, 1970). To test whether the PT-limits were well set, the impact of PT on the recognition process was analyzed. It was expected that the longer the exposure of the target was, the greater the probability of recognition would be. If there is no influence of PT, then the recognition processes are not sensitive to the PT range used here.

Thatcher speed

By turning Thatcher faces upside-down, the local region of the eyes and the mouth is upright due to the specific manipulation of Thatcher faces. Within a *speeded identification task* with limited PTs, several distinguishable hypotheses can be tested concerning the recognition speed.

Firstly, if there are no differences of RTs between original faces and Thatcher faces, then the specific alterations of Thatcher faces are not influential for the processing speed. This means that only the outer area of the faces is involved in the early identification processes within the given time window, because the only differences between the two stimulus classes lie in the inner parts.

Secondly, if shorter RTs reveal an advantage of thatcherised faces, then very early identification processes profit from an analysis of single inner features. In a Thatcher face, the most important inner features (the eyes and the mouth) are in the usual correct upright orientation, so they can be identified without any *mental rotation* (Shepard & Metzler, 1971; Cooper & Shepard, 1973; Cooper, 1976; Lawson & Jolicoeur, 1999)[52]. Mental rotation is a time consuming process, particularly for complex features (Jolicoeur, 1985; Stürzel & Spillmann, 2000). Therefore, the identification of the eyes and mouth region of inverted Thatcher faces might spare time!

Thus, using an inverted Thatcher face paradigm allows to dissociate the processing of global from local facial information. This procedure is a more direct methodical account than procedures in which the spatial scale of the images was varied (e.g., Schyns & Oliva, 1994).

For all RT-analyses, only ‘same’ pairs were used to which the participants responded with “yes”. *Yes-same* trials include responses to original faces as well as to Thatcher faces, whereby original faces have been correctly identified as originals and Thatcher faces have been falsely identified as originals (see section *Material*). Correct rejections (responding with “no” to different trials) were not included, because such responses are rather problematic within this paradigm. These responses include correct rejections of *different* trials as well as rejections of Thatcher faces, which are identified as being thatcherised. However, a rejection of a Thatcher face being identified as thatcherised seems to be a rather different cognitive task than a rejection of non-same identities. Therefore, it is problematic to compute both measures for one single variable. This is not the case with *yes-same* trials, where Thatcher faces have

---

[52] For a different working hypothesis about disoriented faces see Corballis (1988).

not been identified as thatcherised. Further evidence for this argument is given by the empirical data, which support this notion. The oddness of Thatcher faces goes practically undetected and, therefore, they are identified as originals without distinguishing an original face from a Thatcher face. For this reason of being indistinguishable, both measures were aggregated into one dependent variable.

Hypotheses.

In the Experiment 3a, four separate hypotheses were investigated.

- Hyp.1–*Thatcher detection*: There is a generally known effect of inverted Thatcher faces, namely that they are hardly perceived as altered versions, but that they are perceived as well as the original itself. However, it must be tested whether there is indeed no difference between both face classes. The confutation can be demonstrated by a difference of discriminability measures like *A'* and by calculating the *yes-same* rate between original and Thatcher faces.

- Hyp.2–*PT effect*: There is a general recognition advantage for longer PTs. It was predicted that the longer the PT was, the greater the probability of recognition would be. In terms of discriminability, it is predicted that a longer PT will increase *A'*.

- Hyp.3–*Thatcher speed*: In early face processing, local featural recognition processes are particularly important. This yield an advantage for the identification of inverted Thatcher faces over inverted originals. The most prominent local features of inverted Thatcher faces are in the correct orientation. Thus, there is no need for separately turning the main local features of the face into the correct base orientation. Consequently, for early processing (short PT), Thatcher faces are predicted to be identified faster than their original counterparts.

- Hyp.4–*Early processing against later processing*: The advantage for inverted Thatcher faces assumed by Hyp.3 might disappear when a face is presented for a longer time. This might be due to optimized natural face processing strategies, if there is more time to process the face. Prominent candidates for such optimized strategies might be holistic or template-like processes. Therefore, original faces are predicted to be identified faster than Thatcher faces for the long PT.

Method.

Participants. Thirty students participated in the experiment. The participants were undergraduate students (23 women) of the Freie Universität Berlin, who were given a course credit to fulfill course requirements. The mean age was 24.6 years (from 19 to 45 years). All participants had normal vision abilities or were corrected to normal vision. The number of thirty subjects were planned a priori to be able to test several null hypotheses with an acceptable β-failure of .15, which is 3:1 compared to a common set α of .05 (see compromise-strategy in Erdfelder et al., 1996). For all computations, an effect size f^2 was fixed to .35, which is a large effect according to the convention of Cohen (1988). Given these parameters, the power (1-β) for the main effects (df=1,29) was .88 (Erdfelder et al., 1996).

Material. The material was constructed from the frontal photographs of nine female celebrities[53] (from the top of the hair to the bottom of the chin). The celebrities and the specific images were all highly familiar to the participants. To ensure this, subjects were shown all the faces used in the later experiment at the beginning of the session. Only well known faces to which the participant could provide the correct names and some extra semantic information were included in all of the further data analyses. It is essential that the participants were familiar-

---

[53] All celebrities were very popular in Germany, which was investigated in a pretest. The names of the celebrities are given in the description of Figure 4-40.

ized with stimulus material. According to Bruce (1982), there is an important conceptual difference between familiar and unfamiliar faces. The former are already represented structurally and semantically in long-term memory. Therefore, subjects are able to use given primes more efficiently. Moreover, perceptual details like local characteristics and fine configurational information is available without further training. To exclude artifacts of colorization of the pictures, the photographs were transformed into a uniformly *high color format* (2^{16} colors) with 72 pixels per inch, fitting in a graphic window size of 220^2 pixels.

Two different stimulus versions (later called CLASS) of the famous faces were used. First, the *original* class of unmanipulated pictures of the celebrities. The second class consisted of thatcherised faces (shortly called *Thatcher faces*) of the same celebrities. In Thatcher faces, the areas of the eyes and the mouth are turned by 180°. Furthermore, the resulting edges of these areas were smoothed by the image-editing software Adobe Photoshop 4.0 (Adobe Systems Incorporated, 1997) to remove graphical inconsistencies and high degrees of salience of the pictures (like in Leder et al., 2001). An illustration of the original and thatcherised faces is given in Figure 4-40.

Figure 4-40: Face material used in Experiment 3a (inverted faces) and Experiment 3b (upright faces). From left to right: Julia Roberts (actress), Claudia Schiffer (super-model), Lady Diana (royal), Marilyn Monroe (actress), Cindy Crawford (super-model), Verona Feldbusch (national TV-star), Cameron Diaz, Gwyneth Paltrow and Pamela Anderson (actresses). The top row shows the originals as uprights that were *not* used in Experiment 3a, but are shown here for better illustration. In the middle row are the inverted originals and in the bottom row the thatcherised versions are shown.

Procedure. The experiment was conducted on a Macintosh Imac-350 with an integrated 15″–CRT-screen. The resolution of the monitor was 1024 x 768 pixels with a 75 Hz-refresh rate, resulting in an averaged size of the faces of 5.0 cm x 6.5 cm or a visual angle of 2.4° x 2.8° with subjects sitting about 60 cm in front of the screen. The luminance of the screen was 220 cd/m^2 (see for measuring details Experiment 1). The experiment was controlled by the computer program *PsyScope PPC 1.25* (Cohen et al., 1993), which allows the presentation of stimuli within one CRT-refresh cycle. A *CMU ButtonBox* managed the registration of the subject's responses with a time measurement accuracy of one millisecond.

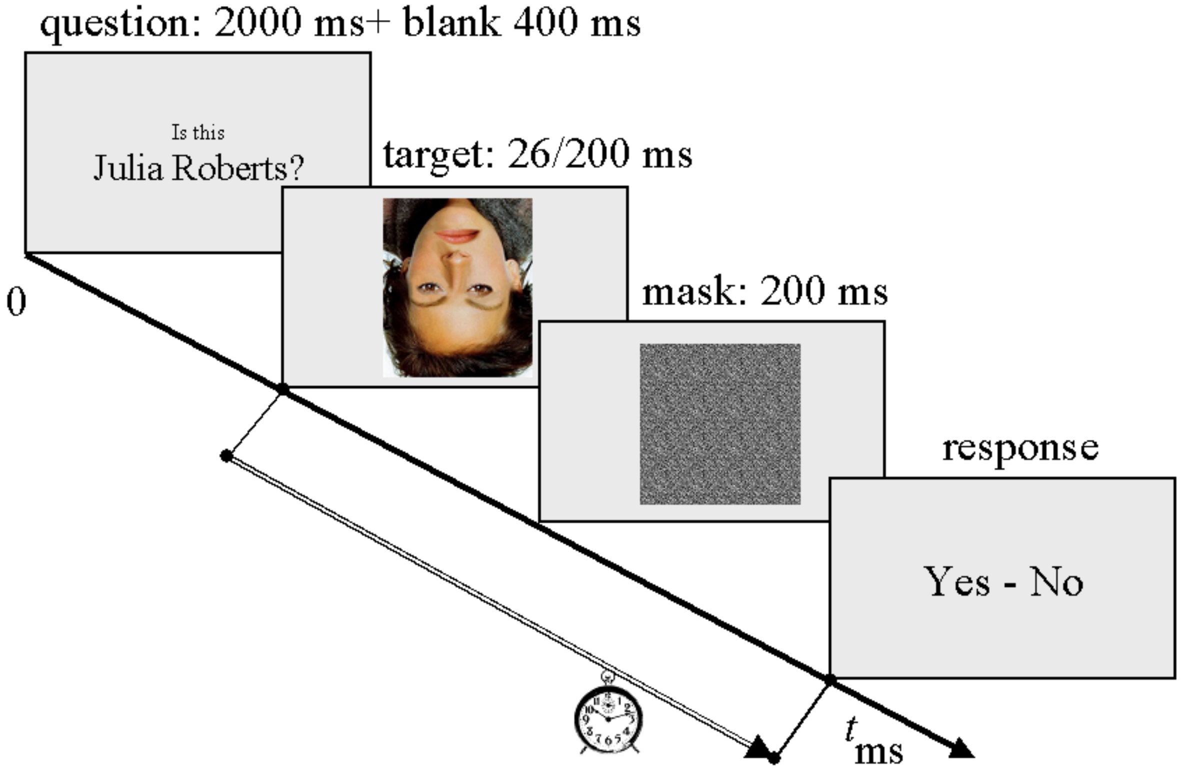
In the beginning of the experiment, all original pictures of the celebrities were presented upright on the screen and the subjects had to name all faces they know very well. Each correctly named face was then faded out, leaving only the unknown pictures back on the screen. The left over faces were then named by the experimenter herself. To realize a well established familiarization of the stimulus material and a subsequent optimal face recognition process (Ellis et al., 1979), unknown pictures were not included in the following statistical analysis.

Immediately after this familiarity check, the test phase began. Each trial commenced with the question: 'Does the following picture show an *original* facial picture of <pre- and surname of one of the nine celebrities>?'. All names belonged to the set of famous faces.

The instructions and questions during the experiment were given in German. A *(name)-verification* task was adopted, because it is a general finding that when people form visual images of things to be remembered, retention is remarkably better than when images are not formed (Bower, 1970; Paivio, 1969). This seems ideal for the procedure of a *speeded decision task*, for which RTs are calculated.

After the initial question (2000 ms), there was a 400 ms lasting blank screen followed by the target, which was the inverted picture of one of the nine celebrities, either in an unmanipulated (original) or a thatcherised version. The participants had to answer as quickly as possible (within a SOA of 5 seconds) and accurately whether the answer to the introductory question was ‘Yes’ (true) or ‘No’ (false). The keys belonging to the answers were alternated over the subjects. In 50% of the cases, the face followed the ‘correct’ name. The time-course of one trial is illustrated in Figure 4-41.

Figure 4-41: Time-course of Experiment 3a for each test trial. In this example, it was asked for Julia Roberts and as the following picture, the thatcherised face of Julia Roberts was presented.



Two presentation times (PTs) were used. Either 26 ms (*short PT*) or 200 ms (*long PT*) due to the refresh rate of the CRT-monitor used (possible PTs were multiples of 1000/75 ms=13.33 ms). The face was followed by a 200 ms random dotted visual mask (same mask as in the experiments before; see illustration in Figure 8-6). All subjects were tested individually.

It is important to note that the question of the requested task was to recognize the *original* face. The participants were explicitly instructed to answer only with ‘Yes’ when they were sure that it was not only the compatible face *but also* that this face was *not altered*! This special kind of question enables us to test whether Thatcher faces can be detected. Thus, it was expected that participants would react with ‘No’ when they had recognized something unusual in ‘same’ trials.

The whole procedure lasted about 30 minutes including instructions and a post-experimental interview, in which the participants were asked whether they had detected some anomaly in the stimuli.

Results & Discussion.

For the following analyses, only familiar faces were included. 93.9% of all faces had been classified as familiar.

For investigating the distinguishability for both face classes (Thatcher vs. Original), *A’* values and *yes-same* rates were calculated (see Table 4-24). The sensitivity measure *A’* is elaborately described in section 4.2.2.1.

Table 4-24: *Yes-same* rates and *A'*-scores (in round brackets) for both face classes (Original vs. Thatcher) and both PTs (long vs. short). The *SD*s of *yes-same* rates are given in square brackets.

	Original			Thatcher		
	yes-same	*SD*	*A'*	*yes-same*	*SD*	*A'*
PT-short (26 ms)	.822	[.120]	(.925)	.739	[.189]	(.904)
PT-long (39 ms)	.905	[.093]	(.960)	.769	[.244]	(.940)
Average	*.864*	*[.114]*	*(.943)*	*.754*	*[.217]*	*(.921)*

Hyp.1–Thatcher detection

In order to test whether (inverted) Thatcher faces and (inverted) original faces are discriminable, the *A'* data of each participant were submitted to a two-way repeated-measurement ANOVA with CLASS and PT as within factors and *A'* as dependent variable (see Table 8-33 in the appendix).

The factor CLASS was not found to be significant ($F_{1,28}$=3.41; P=.0755, n.s.), thus Hyp.1 was not supported. The participants did not identify the originals more often as *originals* than Thatcher versions (see Table 4-24). However, the high *F*-value indicates that there was a tendency for this. However, with an *a priori* test power for large effects of .88, it is very likely that there is no advantage for identification of one of the face classes. This is in accord with other research on Thatcher faces (e.g., Leder et al., 2001; Bartlett & Searcy, 1993). Moreover, none of the subjects detected any odd qualities in the Thatcher faces when queried in a post-experimental interview (cf. Valentine & Bruce, 1985). In line with this finding, Lewis and Johnston (1997) demonstrated that inverted Thatcher faces require a very long time to be identified as Thatcher faces in a matching task.

Nevertheless, using *A'* seems not completely unproblematic. *A'* was used here as based on data gathered from non-distinguishable Thatcher faces and original faces. In common experimental designs, *A'* is an integrative measure of hits and false alarms (see section 4.2.2.1). In the present design, the definition of hits as well as false alarms is rather tricky. It was assumed that people could not detect any oddness in Thatcher faces when they are presented upside-down. However, what could an appropriate measure be for such an experimental design? Hits are defined as correctly identified targets, in this case the identification of original faces in 'same' trials. More difficult are the responses to Thatcher faces.

On the one hand, according to the empirical findings that Thatcher faces are not detectable as being odd, but rather are seen as normal faces, the criterion for correct or false trials could be defined as follows. For a correct response, a participant has to answer with "yes", regardless of whether the stimulus is a Thatcher face or an original one.

On the other hand, according to the specific interrogative form of the task, every trial with a Thatcher face would be correct only if the participant had responded with a "no".

Therefore, the *A'* as it was used here seems *not* to be fully appropriate to test whether there is a recognition difference between Thatcher faces and original faces. Additionally, using *A'* seems to be circular in this design.

Alternatively, the *yes-same* rates (see data in Table 4-24) were analyzed. *Yes-same* rates indicate whether there are any recognition differences between Thatcher faces and original faces. If a participant would respond 100% correctly by categorizing all Thatcher faces and original faces, then the *yes-same* rate would be 0.0 for Thatcher faces and 1.0 for original faces.

A two-way repeated measurement ANOVA with CLASS and PT as within factors and *yes-same* rate as dependent variable was conducted (see further details in Table 8-32). Contrary to the analysis of *A'* above, the main factor CLASS was now significant ($F_{1,29}$=12.561; P=.0014) with higher rates for original faces than for Thatcher faces. No other effect was found to be significant.

Therefore, *yes-same* rates revealed support for Hyp.1. Subjects were able to detect irregularities in Thatcher faces. Nevertheless, they were not able to perform on a high level. As shown in Table 4-24, the averaged *yes-same* rate for Thatcher faces was rather high (*rate*=.754). Therefore, 75.4% of all Thatcher faces were recognized as normal "original" faces.

This high level is advantageous for testing specific processing hypotheses, because only then does it makes sense to compare both classes in more detail (e.g., for their RTs).

Hyp.2–PT effect

Additionally, Hyp.2 was validated by a significant factor PT ($F_{1,28}$=8.378; *P*=.0073) in the *A'*-ANOVA from above. A longer PT (200 ms vs. 26 ms) increased the discriminability *A*' (see Table 4-24).

The first two hypotheses served only as validation for whether the material and the time conditions are well set, but they are an essential premise for the specific testing of processing hypotheses nonetheless.

Hyp.3–Thatcher speed

Hyp.3 and Hyp.4 test specific processing models and are concerned with RTs. Only *yes-same* trials were used. *Yes-same* trials are trials for which the person questioned and the presented picture of the person were compatible and the participants responded by saying "Yes, it is the same".

Table 4-25: RTs (in ms) for *yes-same* targets. The *SD*s are given in square brackets. The last column shows the *Thatcher advantage effect* (TAE) in ms. A description of the TAE will be found below.

RT (*yes-same*)	Original		Thatcher		TAE
PT-short (26 ms)	1130.6	[280.8]	1062.9	[206.9]	67.7 ms
PT-long (200 ms)	997.1	[371.4]	1070.8	[349.7]	-73.7 ms

If Thatcher faces are more easily recognized in a brief presentation, then there must be a significant RT difference for the shorter PT (Hyp.3). Additionally, if this advantage is lost after 200 ms, there must be an interaction between CLASS and PT with a significant difference between Thatcher and original faces under the longer PT (Hyp.4).

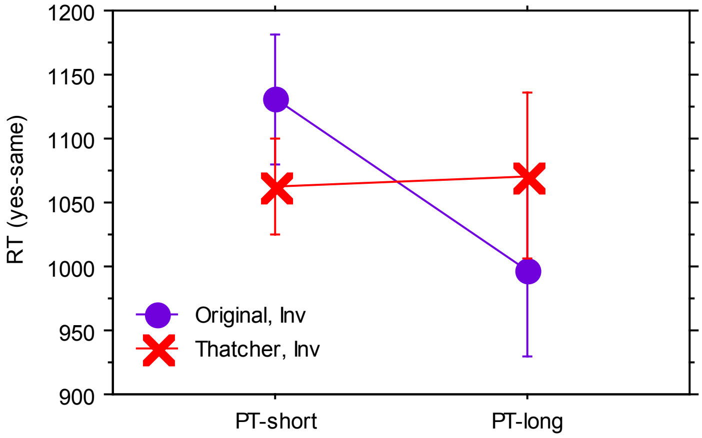
Both hypotheses were tested by a repeated measurement ANOVA with CLASS (Thatcher vs. Original) and PT (short: 26 ms vs. long: 200 ms) as within factors (see Table 8-34 in the appendix). The dependent variable was the RT data for *yes-same* trials.

There was no main effect of CLASS ($F_{1,28}$=.011; *P*=.9179, n.s.) and PT ($F_{1,28}$=1.751; *P*=.1964, n.s.). This means that there was no *general* difference between original and Thatcher faces or between short and long PTs. Nevertheless, an interaction between CLASS and PT ($F_{1,28}$=6.963; *P*=.0134)[54] revealed that the factor CLASS had differential influence under different PTs. As illustrated in Figure 4-42, the RT advantage for Thatcher faces for the short PT condition was reversed for the long PT condition.

---

[54] Due to some missing data of 'Yes' answers to same pairs, not all RT data cells were occupied, so this analysis underwent a reduction of *df*s from initially 29 to 28.

Figure 4-42: Interaction CLASS * PT of RTs for *yes-same* targets. Error-bars are *SE*s.



Moreover, the RT advantage for Thatcher faces under PT=26 ms was found to be significant by a one-tailed *T*-test (T_{29}=2.15, *P*=.0020), which lends support to Hyp.3.
As shown in Table 4-25, this *Thatcher advantage effect* (shortly called TAE) was substantially large (67.7 ms).

Hyp.4–Early processing versus later processing

However, this TAE (67.7 ms) for longer PTs developed into a disadvantage for Thatcher faces as compared to original faces in terms of RT data (-73.7 ms). The negative TAE was also found to be significant (one-tailed *T*-test with T_{28}=2.186, *P*=.0186).
Besides the prominent interaction of CLASS * PT, no other effect was significant as described above, with an *a priori* power of the test of .83. Thus, there were no general differences between both CLASSES over the two PTs.
To ensure that this RT effect was not an artifact of RT-outliers (cf. Snodgrass et al., 1985; Judd & McClelland, 1989), in an additional ANOVA the RT data were restricted to a subject-specific range. In this analysis, the RTs were restricted within ±3 standard deviations (*SD*s)[55] from the individual RT average; otherwise, they were excluded from further analyses. With the same ANOVA design as above, the restricted RT data revealed the same effects (see Table 8-35 in the appendix)[56]. The interaction between CLASS and PT was still the only significant effect ($F_{1,28}$=6.58; *P*=.0160). Again, the RT results show that the nonexistent main effect of CLASS covers up the tremendous underlying change of important face identification strategies when split up in different PTs. The only difference between Thatcher faces and original faces is the area around the eyes and the mouth. Therefore, the reason for the change of a positive TAE[57] to a negative one must be the specific alteration in the face. The experimental strength of the Thatcher paradigm is that variations of the stimuli are attained exclusively through the rotation of the high-level facial features by 180°. Nothing else is changed. Moreover, even the recognition performance is (nearly) the same as shown above—a result

---

[55] The criterion of outliers is not standardized like the α-level. Some authors (e.g. Snodgrass et al., 1985) prefer a criterion of ±4 *SD*s or use a measure of central tendency that is less sensitive to outlying observations such as the median or the geometric mean (see for details section 4.2.2.2). The method of an adaptive criterion to individual RT-pattern is not as strict as a rigidly set absolute limit. No other limits were used. Typically slowed trials after an error (Rabbitt, 1966) were not excluded in favor of a balanced experimental design.

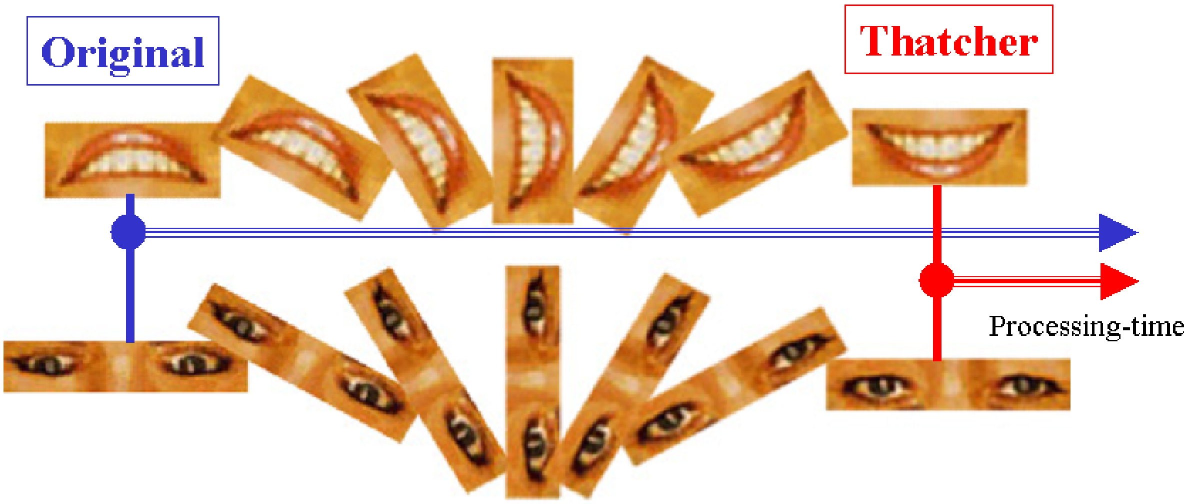
[56] Even an ANOVA with the median as average measure revealed the same results with the only significant effect of the interaction of CLASS * PT ($F_{1,28}$=5.33; *P*=.0286) as well as an ANOVA with RT-criterion of ±4 *SD*s ($F_{1,28}$=6.82; *P*=.0144; see Table 8-36 in the appendix).

[57] The TAE (*Thatcher advantage effect*) is the recognition advantage of (inverted) Thatcher faces over (inverted) original faces in terms of faster RTs.

which is found very seldom in an experimental setting. What might be the reason for Thatcher faces being recognized faster at shorter PTs than the originals themselves?

The result of the Thatcher rotation manipulation is a somewhat paradoxical configuration. For an inverted Thatcher face, the rotated features (eyes and mouth) are in an *upright* orientation. In addition, as it was described in section 2.2.2, our face expertise is mainly limited to upright features and faces. However, this correct orientation comes with a resulting incoherence of the (upright) local features with the rest of (inverted areas of) the face. Due to a significant difference for CLASS under the short PT, this specific rotation must be advantageous for the identification of (inverted) faces. The human face identification is optimally suited for upright faces, but for inverted Thatcher faces only the manipulated local parts are recognized optimally. A TAE for briefly presented faces indicates that within a short PT, especially the identification of local features is advantageous for the recognition of the whole face. The disadvantage of an incoherent facial outlook does not seem to be an impediment. Rather, the just-in-position of the local features need no further rotation process—and this saves time! Figure 4-43 illustrates this with the eyes and mouth area of Lady Diana.

Figure 4-43: Illustration of the *mental rotation hypothesis* with facial parts of the face of Lady Diana. According to the mental rotation hypothesis, an object has to be rotated to the commonly used orientation, which requires additional processing-time. In this example, the main facial parts of an (inverted) original face are presented upside-down. Therefore, they have to be rotated. In contrast, the facial parts of an alternative Thatcher version are already in the common orientation. Therefore, the time required to process these features is shorter (symbolized by a red arrow) than the time needed to process those of the original face (blue arrows).



This result fits very well into the growing literature on mental rotation. It has often been demonstrated that there is a stable RT pattern with (linear) increasing RTs by the rotation of objects from their familiar viewpoint (Cooper & Shepard, 1973; Cooper, 1976; Shepard & Metzler, 1971; Jolicoeur, 1985; Tarr & Bülthoff, 1995). Therefore, the identification of local features seems to be a substantial part of the face processing in the first milliseconds!

Interestingly, Tarr et al. (1998), who studied the recognition of rendered single geons from Biederman and Gerhardstein's (1993) Experiment 4 with line drawings, found that rotation rates for such stimuli ranged from approximately 750°/s for a naming task to 3600°/s for a match-to-sample task with a Go/No-Go response. This corresponds to a total time of 50 ms–

240 ms for a full 180° rotation[58], which is approximately in the range of the Thatcher advantage effect demonstrated above (see Table 4-25).

Nevertheless, the advantage of early local processing does not rule out the possibility that at the same time more global processes may already be in progress, for instance the processing of the outline (hair, chin) or the skin texture. Indications for the involvement of such processing were found in the post-experimental interviews with many participants, who reported that they sometimes had only seen outlines or contours of some faces. Additional empirical results came from Phillips (1979), who found that inversion affects recognition of the internal features of the face (eyes, nose, mouth) more than recognition of the external features (hair, chin). Moreover, in a matching study of Young et al. (1985, Exp.2), external feature matches were obtained much faster than were internal feature matches. It was further proposed that global configurations are crucial in activating stored object representations (Boucart, Humphreys, & Lorenceau, 1995).

The present experiment cannot rule out a global precedence, because only a further reduction of PT could have the power to reveal the nature of this supposed outline identification process. Another possibility would be to test the RT data for faces with exclusively global information in comparison to normal faces (this will be done in Experiment 4 and Experiment 5). Therefore, for Experiment 3a it remains unsolved whether global processes are occurring simultaneously with the local ones or whether one of the processes precedes the other in a kind of dependence. Nevertheless, it was shown that local identification processing is already established after a very short PT of only 26 ms.

Furthermore, the benefit of early local feature analysis seems to be greater than the cost of failed coherence between the local parts and the disoriented whole. In other words, during an early stage of processing there is actually no possibility to detect such an incoherence, because the different processing levels (local and more global ones) are not yet binded together (see section 3.1.3 and section 3.2 for the *binding problem*).

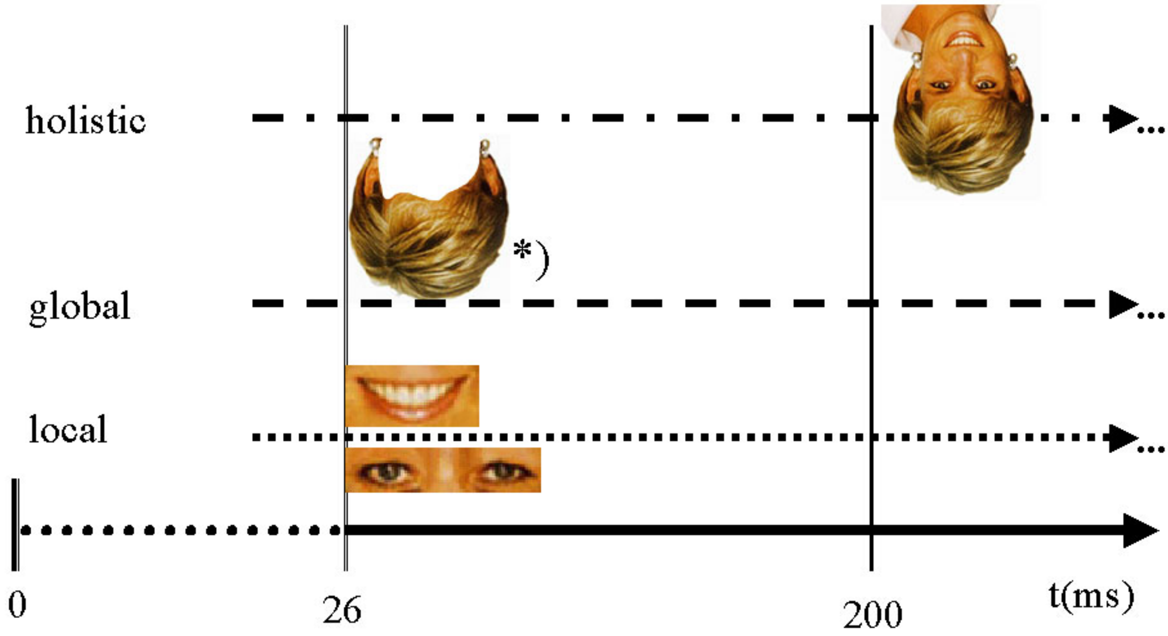
The RT results for the long PT indicate that this binding was also not possible later on: the identification speed of Thatcher faces was not impeded by a long PT (two-tailed T-test with T_{28}=.30; P=.6169, n.s.), but the local identification *advantage* was lost. This can be seen in the interaction of CLASS * PT. The ‘Thatcher advantage effect’ changed to an ‘Original advantage effect’. The RTs for the long PT were shorter for original faces than for the Thatcher faces as analyzed above.

It was not tested in this experiment which specific advantageous processes come into play for normal (original) faces under the long PT, but these processes seem to be optimized for normal faces. Most probably, template-like matching routines compare the face with stored templates. The results from Experiment 3a are combined in a second face-processing model called FACEREC-2, which is illustrated in Figure 4-44.

---

[58] This projection, of course, must not be taken literally. From a view-based perspective, a rotation of 180° would be expected to produce enormous rotation costs, relative to slight rotations angles. The opposite occurs, however (Biederman et al., 1999): mirror reflections incur no cost in priming (Biederman & Cooper, 1991) or less effort than angles of about 90° for some other tasks (Valentine & Bruce, 1988, Figure 5).

Figure 4-44: Face processing model FACEREC-2 in line with the results from Experiment 3a. Processes which are particularly advantageous for the recognition of faces are split in three groups. A holistic (whole face), a global (global structures, outline), and a local one (face parts: eyes, mouth). Local feature identification is particularly beneficial for the identification of briefly presented faces. The evidence for global processing, indicated by a '*' is based only on qualitative data reported by the participants in post-experimental interviews.



This simple FACEREC-2 model combines the results of Experiment 3a. It is most obvious that local features are identified very early. A subsequent coherence check or holistic processing starts later.

For the Thatcher faces used in this experiment, it is, of course, not clear whether the mouth *or/and* the eyes are responsible for the TAE, because both areas were always changed concurrently. From data of Experiment 1, the eyes region seems to be the more prominent candidate, but also the upright mouth with its high physical distinctiveness and its high amount of meaningful social information (Parks et al., 1985) seems plausible. It is further assumed that global structures are recognized simultaneously. Evidence for this hypothesis, however, only comes from post-experimental interviews with the participants (indicated in Figure 4-44 by a '*'-symbol). Moreover, other researchers have argued that the visual system does not need more time to additionally process the lower resolution image (Lakshminarayanan et al., 1997).

In order to systematically test whether global structures are already available within a very brief PT and to control whether coherence of local and global information is checked within such limited time resources, the next experiments were planned.

4.4.2 Experiment 3b: Upright Thatcher faces

Introduction.

The most important aim of this experiment was to provide a control condition for Experiment 3a. While Experiment 3a used inverted faces as a control condition, upright faces had to be used here.

Furthermore, it was investigated whether upright Thatcher faces are also identified as destructed faces even in a very early stage of processing or whether this identification performance is only mediated through later processes. This allows a better understanding of whether any *binding* of internal (local) and external (global) structures is possible within the first milliseconds of face processing. Several hypotheses have to be tested.

Thatcher effect

It is predicted that upright Thatcher faces are easily recognized as being distorted. This will be tested by two analyses. On the one hand, the discrimination performance (*A'*) of Thatcher faces with original faces will be investigated. Following the discussion of the problematic use of *A'* in Experiment 3a, *yes*-same rates will be analyzed additionally. According to Thompson's (1980) Thatcher face demonstration, upright faces should be detected very accurately as being odd, even after a very brief presentation. On the other hand, referring to Experiment 3a, upright Thatcher faces should be identified as being distorted more often than inverted Thatcher faces.

Upright superiority

Face processing is highly optimized for upright faces (see section 2.2.1). Therefore, it is predicted that upright original faces compared with inverted original faces should be recognized better and faster in terms of *yes-same* rates and RT data, respectively.

Inhibitory effect

Moreover, upright Thatcher faces are more difficult to identify due to their grotesqueness, and they are never viewed as an 'original' image of a person. In addition, the inner and outer parts are incoherent, because of the inverted eyes and mouth region. This should result in longer RTs than for original uprights.

Method.

Participants. In order to compare Experiment 3a with 3b, the same number of students participated in the experiment. The 30 participants were undergraduate students (20 women) of the Freie Universität Berlin, who were given a course credit to fulfill course requirements. The mean age was 26.1 years (from 19 to 39 years). None of the participants had taken part in Experiment 3a.

Material. As material, the same stimuli as in Experiment 3a were used. In Experiment 3b, they were presented in an upright orientation.

Procedure. Pre-studies revealed that even under short PTs, subjects were able to notice inconsistencies of distorted upright faces—in this case upright Thatcher faces. Based on these results, the long PT condition of Experiment 3a was reduced in Experiment 3b from 200 ms to only 39 ms. This was done to prevent floor effects of the 'Yes'-rate of thatcherised faces. The other procedural parameters were the same as in Experiment 3a.

Hypotheses.

Experiment 3b explicitly addressed three different hypotheses.

Hyp.1–*Thatcher effect:* Upright Thatcher faces are easily recognized as being distorted. It is predicted that Thatcher faces are not only accurately recognized as being odd but are also much more accurately recognized than inverted Thatcher faces from Experiment 3a.

Hyp.2–*Upright superiority*: Face processing is highly optimized for upright faces. It is predicted that upright original faces compared to inverted original faces should be recognized better and faster in terms of *yes-same* rates and RT data, respectively.

Hyp.3–*Inhibitory effect*: Upright Thatcher faces are perceived as grotesque and seem to operate as highly emotional cues. Moreover, the inner and outer facial parts are not in coherence. Therefore, an inhibitory effect on the identification speed (in terms of longer RTs) is expected, because they are processable less fluently than undistorted faces.

Results & Discussion.

Only familiar faces were included in the subsequent analyses. 91.8% of all faces were recognized as being familiar.

Table 4-26 shows the mean *yes-same* rates for both face classes (Original vs. Thatcher) and for both PTs (short: 26 ms vs. long: 39 ms). Additionally, the *A'* values are given in round brackets.

Table 4-26: Identification rates for same Targets. The main data shows the *yes-same* rate, whereas the *SD*s are presented in square brackets and the *A'* values in round brackets.

	Original			Thatcher		
	yes-same	*SD*	*A'*	*yes-same*	*SD*	*A'*
PT-short (26 ms)	.927	[.083]	(.970)	.528	[.322]	(.871)
PT-long (39 ms)	.935	[.100]	(.964)	.503	[.341]	(.861)
Average	*.931*	*[.091]*	*(.967)*	*.516*	*[.329]*	*(.866)*

Hyp.1–Thatcher effect

In order to test Hyp.1, the *yes-same* rates and the *A'* values were analyzed. The extent of the detection of inconsistencies in faces can be tested with the indirect measure of *yes-same* answers, because the participants were asked to give only positive ('same') answers when the target shown was exactly the same as the original face known from TV, magazines, or movies. Thus, if the yes-rate drops, this is an indication of correctly rejected Thatcher faces. The *yes-rates* for same pairs were submitted to a two-way repeated measurement ANOVA with CLASS (Original vs. Thatcher faces) and PT (short: 26 ms, long: 39 ms) as within subjects factors (see Table 8-38 in the appendix). There was a huge difference between thatcherised faces (rate=.516) compared with original faces (rate=.931), revealed as being significant by the main effect CLASS ($F_{1,29}$=45.45; P<.0001). No other effect was significant. The CLASS effect and the lack of an interaction of CLASS and PT are both clear evidence that upright thatcherised faces were already detected after 26 ms, without any additive benefit over the next 13 ms. This supports Hyp.1. However, the data also reveal that participants were not very sure whether or not the stimuli of this CLASS were distorted. An indication for this might be the relatively high *SD*s for the *yes-same* rate of Thatcher faces seen in Table 4-26.

It is possible that the participants detected the oddness of Thatcher faces, but perhaps had only a feeling that 'something' was wrong with these faces without knowing what this was exactly. Some participants reported after the experiment that they had seen some strange looking faces, but did not know in what respect they were 'strange'. Although these reports were vague, they qualitatively support the assumption that the early identification processes were only rudimentary and not fully conscious.

More importantly, the data reveal that even after a brief period of 26 ms, it was possible that an upright Thatcher face was identified as being odd. This means that the cognitive apparatus is capable of testing an upright face within a very brief time for consistency. Thus, inner facial information is already taken into account for such a short PT. The high *A'* values (no *A'* was below .887; see Table 4-26) of both face classes corroborate this result, although the *A'*-values were significantly lower for Thatcher faces than for original faces ($F_{1,27}$=40.70, P<.0001; see full ANOVA results in Table 8-37 in the appendix), probably due to a higher response uncertainty for thatcherised versions[59].

---

[59] The use of the measure *A'* is rather problematic in this specific experimental design. See Experiment 3a for more details on this topic.

To further analyze the Thatcher effect in general and the time needs of Thatcher faces in particular, the RT data of *yes-same* pairs were investigated in the following tests. Table 4-27 shows the RT data.

Table 4-27: RTs (in ms) for *yes-same* targets. The *SD*s are given in square brackets. The last column shows the difference between the RT of thatcherised and original versions in ms.

RT (*yes-same*)	Original		Thatcher		TAE
PT-short (26 ms)	1089.9	[343.0]	1395.6	[434.5]	-305.7 ms
PT-long (39 ms)	1030.4	[334.0]	1220.8	[525.9]	-190.4 ms
average	*1060.1*		*1308.2*		*-248.1 ms*

The RT data (for *yes-same* trials) were submitted to a repeated measurement ANOVA with two within factors: CLASS and PT. The significant main effect of CLASS ($F_{1,27}$=12.65; P=.0014) revealed a substantial difference between RTs of Thatcher and original faces of 248.1 ms (1308.2 ms vs. 1060.1 ms; see split data in Table 4-27). The results of this RT analysis were very similar to those of the *yes-same* rates discussed above.

The data sampled in the *yes-same* measure include cases in which 'original' faces were correctly identified as being of a famous person plus cases in which Thatcher faces were misinterpreted as being normal originals. Consequently, the values cannot simply be classified as being 'correct', because of this rather unusual task in which a face can be seen as being of a famous person despite the Thatcher manipulation. It is important to note that RT data for thatcherised *yes-same* trials were the RTs of undetected Thatcher faces. It seems that the participants were not only registering the outline of the faces, which were the same in all cases. This can be ruled out because in this case the identification speed would have been the same as the identification speed for original faces. This is not the case, however, which is revealed by the averaged difference of 248.1 ms between Thatcher faces and original faces (see Table 4-27). Therefore, even for the misclassified Thatcher faces, which were found not to be original faces, a coherence check of inner and outer facial parts had been established.

Such an indication was not found for inverted faces, as investigated in Experiment 3a. In Experiment 3a, the availability of the processing of inner facial features had been demonstrated for upside-down faces. Moreover, a parallel ongoing identification of the outer space had been assumed. However, there was no sign for an integration of these two different processes. For upright faces, this seems to be quite different. Here, the inverted eyes and mouth region interfere with the processing of the whole.

Thus, it has to be assumed that higher-order and more integrative early processing is involved in the face recognition of briefly presented upright faces, in comparison to the recognition of inverted ones (cf. paragraph 4.4.1).

Hyp.2–Upright superiority

In order to test the differences between upright and inverted faces statistically, two additional overall ANOVAs, combining Experiment 3a and 3b, were conducted. On the one hand, it is important to compare the detectability of Thatcher faces by the *A'* measure. On the other hand, the processing was investigated by overall RT data.

Both analyses were conducted as repeated measurement ANOVAs with the within factors CLASS (Thatcher vs. Original) and PT (short vs. long)[60]. In order to be able to compare both experiments, an additional between-subjects factor ORIENTATION (inverted: Exp.3a vs.

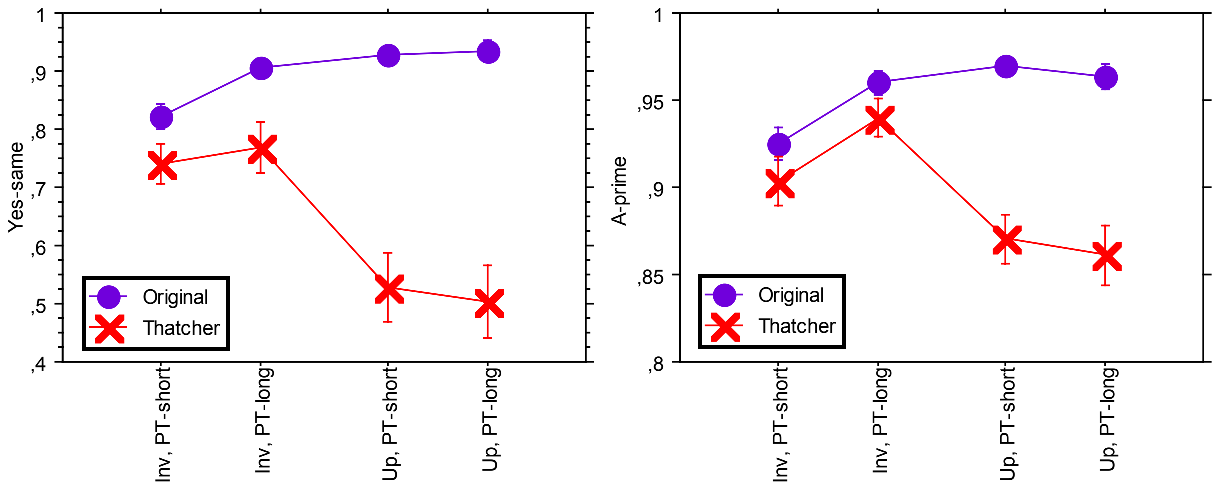
---

[60] The inclusion of factor PT is rather problematic, because the long PT was operationalized differently. In Experiment 3a it was set to 200 ms, whereas it was set to 39 ms in Experiment 3b. Nevertheless, it could be assumed that the dissociation would have been even stronger if the PT in Experiment 3b had been increased further.

upright: Exp.3b) was used. *Yes-same* rate and RT (*yes-same* trials), respectively, were the dependent variables (see Table 8-40 and Table 8-42 in the appendix).

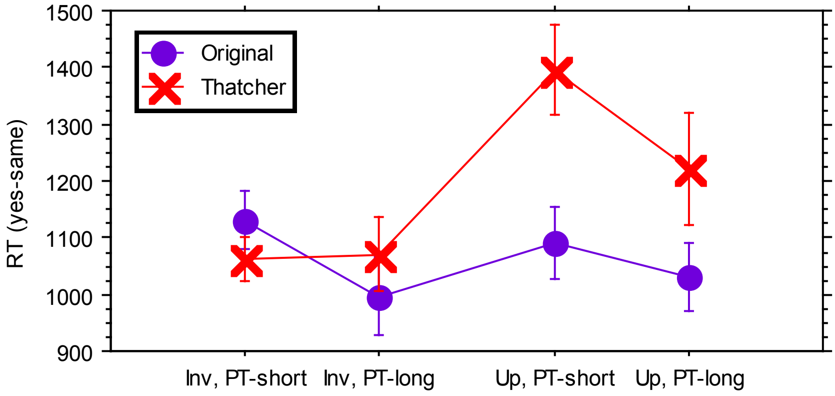
As predicted by Hyp.2, there must be a dissociation between both experiments concerning the different CLASSes. The predicted interaction of CLASS * ORIENTATION was indeed significant in both ANOVAs (*Yes-same*: $F_{1,58}=19.71$, $P<.0001$; RT: $F_{1,55}=11.83$, $P=.0011$). These results generally support a dissociation between upright and inverted faces. As shown in Figure 4-45, there is a dramatic decrease of face recognition sensitivity (measured by *A'*) and *yes-same* rate for upright Thatcher faces in comparison to their inverted counterparts.

Figure 4-45: *Yes-same* rates and *A'* values for both upright and inverted faces, split by CLASS * PT. The PT-conditions are the following: PT-short for both experiments 26 ms, but PT-long for uprights 39 ms and for inverted faces 200 ms. On the left, *yes-same* rates are given; on the right, the *A'* values are shown.



This dissociation pattern is also found in the RT data shown in Figure 4-46. Upright Thatcher faces need much more time to be processed (even if the RT for *no-same* pairs were used as the data to be compared, as demonstrated by the data in Table 8-43 in the appendix).

Figure 4-46: RT (*yes-same*) values for both upright and inverted faces, split by CLASS * PT. For further details on PT-conditions see description of Figure 4-45.y



Moreover, analyzing the RT data revealed another important point concerning *original faces only*. Whereas the discriminatory performance (*A'*) increased for upright faces, there was no difference between the latency data of upright and inverted faces. This would mean that the *speed* of processing original faces was independent of their orientation.

In order to test this, two unpaired *T*-tests (two-tailed) were conducted with *yes-same* rate and RT data of *yes-same* trials. Only the data for original faces in condition PT-short (26 ms) were used, because only PT-short was identical for both experimental designs. The difference between *yes-same* rate scores of inverted and upright faces was significant (T_{58}=3.96, *P*=.0002); but RT-data failed to reach significance between both orientations (T_{58}=.50, *P*=.6169, n.s.). Therefore, on the one hand the *yes-same* rate revealed that there are optimized processes available for upright faces. On the other hand, the processing time was not shorter for upright faces. Thus, only the *yes-same* rate lends support to Hyp.2.

Hyp.3–Inhibitory effect

Contrary to the original faces, the RT data (for *yes-same* trials) of Thatcher faces seem to be highly sensitive to orientation (see Figure 4-46). This was analyzed by an unpaired two-tailed *T*-test for Thatcher faces only. The difference between upright and inverted Thatcher faces for only PT-short was significant (*mean diff*=332.7 ms; T_{57}=3.78, *P*=.0004). There are several possible reasons for the obvious increase in RTs for upright Thatcher faces compared to inverted Thatcher faces.

One explanation might be the high sensitivity for configural alterations in upright faces. Consequently, upright faces would be processed more elaborately. The deeper elaboration needs more time. Therefore, the RT will increase.

Another explanation is based on the finding that Thatcher faces look grotesque (Stürzel & Spillmann, 2000), or that Thatcher faces might be interpreted as affective cues. The interpretation of an affective quality would be plausible for very brief presentations, where it is assumed that recognition operates only on a relatively coarse level. A misinterpretation of an odd face as being an affective face based on low cue quality seems reasonable.

Nevertheless, the latencies for identifying upright Thatcher faces were rather long. However, spending a lot of time to identify emotional aspects of a face might instead be an adaptational disadvantage. Usually, expressive signals are processed quickly and efficiently (Schweinberger & Soukup, 1998). Many authors have emphasized this point and concluded that high emotional cues are processed much faster than neutral ones (e.g. Hogarth, 1987). Similar results indicated that a positive emotion and a highly activated emotion such as surprise or fear are easily recognizable with a relatively brief exposure to the stimuli of 4-64 ms (Ogawa & Suzuki, 1999). Such a fast processing helps to identify rather dangerous but also very positive situations much better and seems to be adaptive and useful from a biological standpoint. Therefore, the RT increase for upright Thatcher faces is not compatible with the common view of the efficiency of processing affective signals. Moreover, why should the identification rate of upright Thatcher faces drop substantially when only emotional states have been registered? It does not make sense that the identification performance drops nearly to chance level when these alterations are recognized only as emotional states and not as new entities.

A last argument against the assumption that Thatcher faces were recognized as affective cues concerns the current view that expression and identity analysis proceed rather independently (Rhodes, 1985; Bruce & Young, 1986; Young et al., 1986)[61], and, therefore, expressions might not have an influence on the speed of identification.

Considering these findings, it does not seem plausible to assume that the cognitive apparatus perceives briefly presented Thatcher faces as being affective cues. Rather, it seems that the high sensitivity for processing upright faces extends the RTs, because incoherence of inner and outer facial parts is detected, and this additional check requires time.

---

[61] Bruce has critically discussed this argument. Namely, such independent processes might share a common early processing stage and, therefore, expressive faces have to be normalized before they can be identified (Valentine & Bruce, 1986b). Part-evidences for such a mechanism can be found in very early works of Bruce. For example, unchanged faces were recognized more quickly and accurately than faces with a change in angle or expression (Bruce, 1982). Furthermore, recent work of Schweinberger et al. (1998) found asymmetric dependencies between different components of face perception. Identity seems to be independent of, but may exert an influence on, expression and facial speech analysis.

4.4.3 Experiment 4: Inverted Thatcher faces with a Gauss mask

Introduction.

The aim of Experiment 4 was to validate findings of Experiment 3a with an alternative matching procedure. There were several points to be investigated.

Nature of the visual mask; Stability of the Thatcher advantage effect (TAE)

First, it is important to check whether the TAE that was found in Experiment 3a is stable if an alternative visual mask is used. Visual masks differ in their effectiveness of deleting the *iconic memory* (cf. Eriksen, 1980; Enns & Di Lollo, 2001). A random dot mask as used in Experiment 3a does not contain facial elements. Therefore, it is assumed that such a mask is not very effective in backward masking of a face stimulus. Thus, an alternative kind of visual mask was used to validate the findings of a *Thatcher advantage effect* (TAE) in Experiment 3a (see Breitmeyer, 1984). See section *Material* for further details.

Enhanced TAE

The second goal of this experiment was to investigate whether the TAE could be increased further by omitting global facial information by cutting the outline of the face. Ellis, Shepherd, and Davies (1979) have demonstrated that the inner parts of familiar faces are more important for recognition than are the external parts. Moreover, Leder and Bruce (1998) showed that effects of distinctiveness based on manipulations of local or configural internal facial features are much stronger when the outer features are omitted from the stimuli. Furthermore, Bruce, Burton, and Dench (1994) found that the distinctiveness of faces with concealed hair (with a swimming cap) correlated highly with the configural attributes of the internal face parts. Therefore, an even stronger TAE is predicted for faces consisting of internal parts only.

Global precedence

In the preceding experiments, local feature analysis and possible configural recognition processes were investigated. Nevertheless, there was only speculation regarding the processing of global structures. In this experiment, a special presentation condition called *OUT*-condition (see for an illustration Figure 4-47), was implemented to test the effectiveness of early global processing. In the *OUT*-condition, only outlines of faces were presented. Global structures like the hair region are coarse and possess a relatively low informational content. Therefore, they are predicted to be processed faster than more detailed structures (Hughes et al., 1984; Paquet & Merikle, 1984).

Rotation

Moreover, it will be tested whether the explanation for the TAE given in Experiment 3a and Experiment 3b is plausible. In the discussion of Experiment 3a, it was assumed that the RT advantage was caused by beneficial early local feature analyses. If it were the case that the processing of local features is advantageous for early face recognition, then this advantage should be decreased by rotating the stimulus material by 45°. What happens when stimuli are rotated? In this case, not only the local features of original faces but also those of Thatcher faces have to be rotated mentally according to the *mental rotation hypothesis* (Shepard & Metzler, 1971). This should decrease the overall TAE.

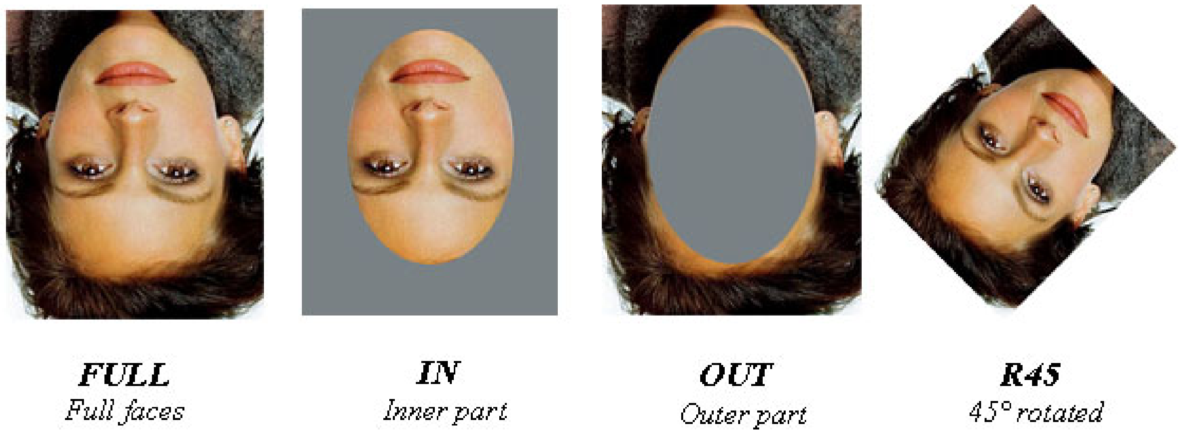
Method.

Participants. In order to be able to compare Experiment 3a with Experiment 4, thirty students again participated in the experiment. The participants were undergraduate students (25 women) of the Freie Universität Berlin, who were given a course credit to fulfill course requirements. The mean age was 25.9 years (from 19 to 42 years).

Material.

The material based on the stimuli of Experiment 3a, in which original faces as well as Thatcher faces were used. There were four different stimulus categories used in Experiment 4. Figure 4-47 and the following description give an overview of this material.

Figure 4-47: Face presentation versions used in Experiment 4 (only shown for original faces, not for Thatcher faces). The face of Julia Roberts was used as an example. From left to right: The unmanipulated face from test block 1 (*FULL*), then typical inner parts presented in block 2 (*IN*), then only the outer part (*OUT*) and lastly the 45° rotated version (*R45*).



FULL-condition

The basic material came from Experiment 3a and was used in the first block of this experiment. For the other experimental blocks, the kind of presentation for these faces was changed. Three additional conditions were used besides the original *FULL*-condition, in which full faces had been presented (called *IN, OUT,* and *R45*-condition).

IN-condition

For the *IN*-condition, only the inner parts of the faces were presented. This was achieved by a standardized oval, which overlays the basic faces. The advantage of using standardized ovals against the commonly used method of sharply cutting the hair is that there is no information available anymore about the form and *Gestalt* of the contours and the hairstyle (cf. Liu & Chaudhuri, 1998).

OUT-condition

In the *OUT*-condition, just the opposite of the *IN*-manipulation was realized. Here, only that part excluded by the *IN*-condition was now apparent.

R45-condition

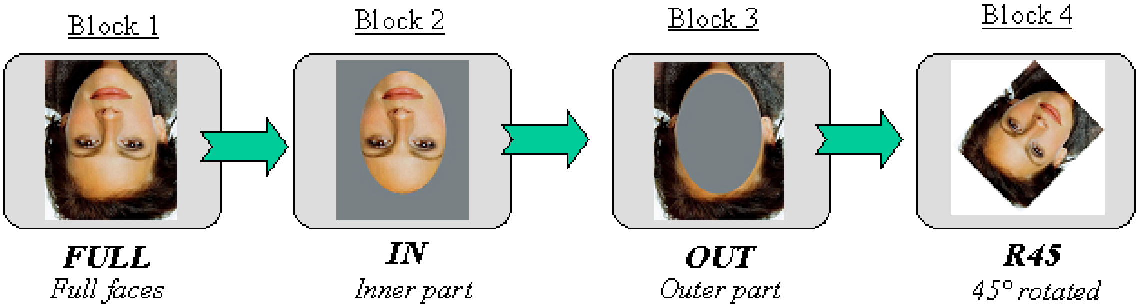
Additionally, in the *R45*-condition the basic faces of the *FULL*-version were presented in a 45°-rotated fashion. Figure 4-47 gives an overview of all presented versions.

Alternative visual mask

In contrast to Experiment 3a, an alternative visual mask was used. Following the suggestions for constructing a visual mask (e.g., Sperling, 1963; Haber & Hershenson, 1973), the random dot pattern used for all preceding experiments was replaced by a composition of all faces used in the experiments, filtered by a Gauss noise. Therefore all nine different face pictures were covered by an image-editing program (Adobe Systems Incorporated, 1997) and afterwards filtered by the same computer program with a Gauss smoothing algorithm (r=300 pixel). See an illustration of this mask in Figure 8-7 in the appendix.

Procedure. The time course of the trials was the same as in Experiment 3a, with the only difference of presenting an alternative masking as described above. Thus, the same PTs and interrogative form was used.
Within the experiment there were four different blocks (referred to henceforth as TASKS) separated by brief breaks, during which specific instructions about the nature of the stimuli were given. Figure 4-48 shows an illustration of the order of blocks. The first block (block1: *FULL*) dealt with full faces, the same as in Experiment 3a. Then followed a block with *IN*-faces (block2: *IN*), after that came a block with *OUT*-faces (block3: *OUT*) and lastly a block with *R45*-faces was presented (block4: *R45*). See the section *Material* for details on the material of these blocks.

Figure 4-48: Sequence of test blocks in Experiment 4. There are four different TASKs with different stimulus categories. See text for details.



Only the *OUT*-condition deviated from the common logic of the blocks. Due to the omission of the inner features in *OUT*-faces, the stimuli did not differ between Thatcher and Original faces. Thus, *OUT*-faces did not belong to a special face CLASS. Each block contained the same number of trials: three preceding blindtrials were followed by 72 test-trials. The whole experiment consisted of 4 x 72 = 288 total test-trials and lasted about 45 minutes. Afterwards, the participants were interviewed about their perceptual experiences during the experiment.

Hypotheses.

In Experiment 4, four hypotheses were addressed explicitly.

Hyp.1–*Stable Thatcher advantage effect (TAE):* The TAE is a stable and strong effect. Therefore, even with an alternative and probably more effective visual mask, the RT for recognizing Thatcher faces is shorter than for recognizing original faces (under the PT-short condition).

Hyp.2–*Enhanced TAE: IN*-faces only contain inner features and micro-level configurations. Therefore, it is assumed that it is very hard to recognize *IN*-faces within a very short time (Liu & Chaudhuri, 1998). If local features are most significant besides global structures under short PTs, then the TAE must be even bigger for *IN*-faces compared to *FULL*-faces.

Hyp.3–*Global precedence:* If only global structures of a face are available (*OUT*-condition), it is predicted that the responses are faster than responses to any other TASK-condition due to their low informational content (Hughes et al., 1984; Paquet & Merikle, 1984). Moreover, according to this logic, *FULL*-faces are assumed to be processed faster than *IN*-faces, because they contain more global structures than *IN*-faces. Most prominently, this might be the case for the short PT condition.

Hyp.4–*Rotation:* It is assumed that the TAE is based on the obligatory rotation of components for inverted faces, which is not needed for Thatcher faces. According to this *mental rotation hypothesis* (Shepard & Metzler, 1971), the processing of a face that is rotated 45° away from the inversion condition can be explained as follows. On the one hand, the eyes and mouth region of normal inverted faces (‘original faces’) need not to be rotated by a full 180° but only by 135°. This should result in faster RTs compared to the fully inverted version. On the other hand, the face-components of Thatcher faces are now in a 45° rotated orientation. Thus, the local features now have to be rotated by 45°. This should result in an RT disadvantage compared to the fully inverted version. The relative advantage of original faces in combination with the relative disadvantage for Thatcher faces should therefore decrease the TAE.

Results & Discussion.

As in the preceding experiments, only familiar faces (90.7% of all faces) were included in the following analyses.

The following analyses were based on the RT data. In order to exclude RT outliers, the RTs were limited to the RT range of ±3 *SD*s around the individual subject's RT average. Table 4-28 shows the RT data for the conditions *FULL, IN,* and *R45* for original as well as for Thatcher faces (factor CLASS). For *OUT*-faces only one RT mean is given due to the non-existing factor CLASS (for an explanation see above).

Table 4-28: RT-data for *yes-same* trials (in ms). *SD*s are given in square brackets. RTs were restricted to the criterion of ±3 *SD*s.

TASK	*PT*	*Original*	*SD*	*Thatcher*	*SD*	*TAE*
FULL	short	1097.8	[349.4]	1093.9	[326.7]	3.9
	long	1069.3	[326.4]	1016.9	[273.1]	52.4
IN	short	1292.4	[441.6]	1284.8	[415.6]	7.6
	long	1128.4	[302.7]	1375.6	[612.1]	-247.2
R45	short	899.4	[256.4]	886.5	[268.2]	12.9
	long	851.6	[227.5]	859.8	[195.7]	-8.2

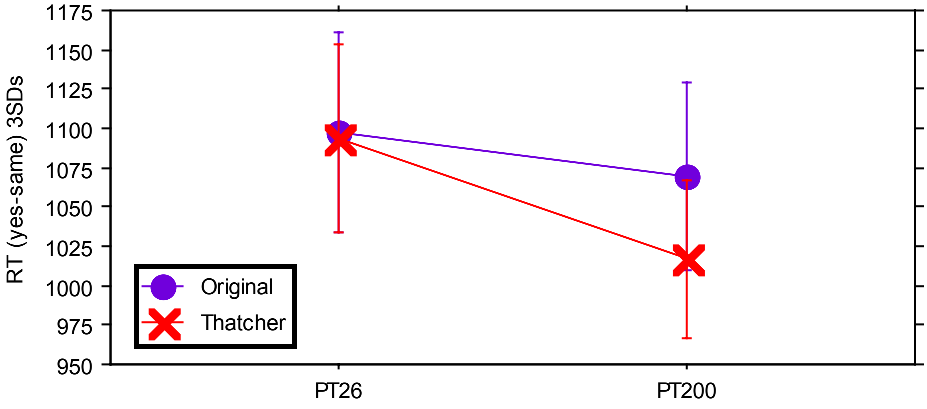
OUT	*PT*	*OUT-faces*	*SD*
	short	967.4	[383.1]
	long	883.5	[260.6]

Hyp.1–Stable Thatcher advantage effect (TAE)

In order to test Hyp.1, a repeated measurement two-way ANOVA with the factor CLASS (Original vs. Thatcher) and the factor PT (short: 26 ms vs. long: 200 ms) was conducted. Only RT data from the first block (*FULL*) with the above-described criterion of ±3 *SD*s were included in this analysis (see Table 8-44 in the appendix).

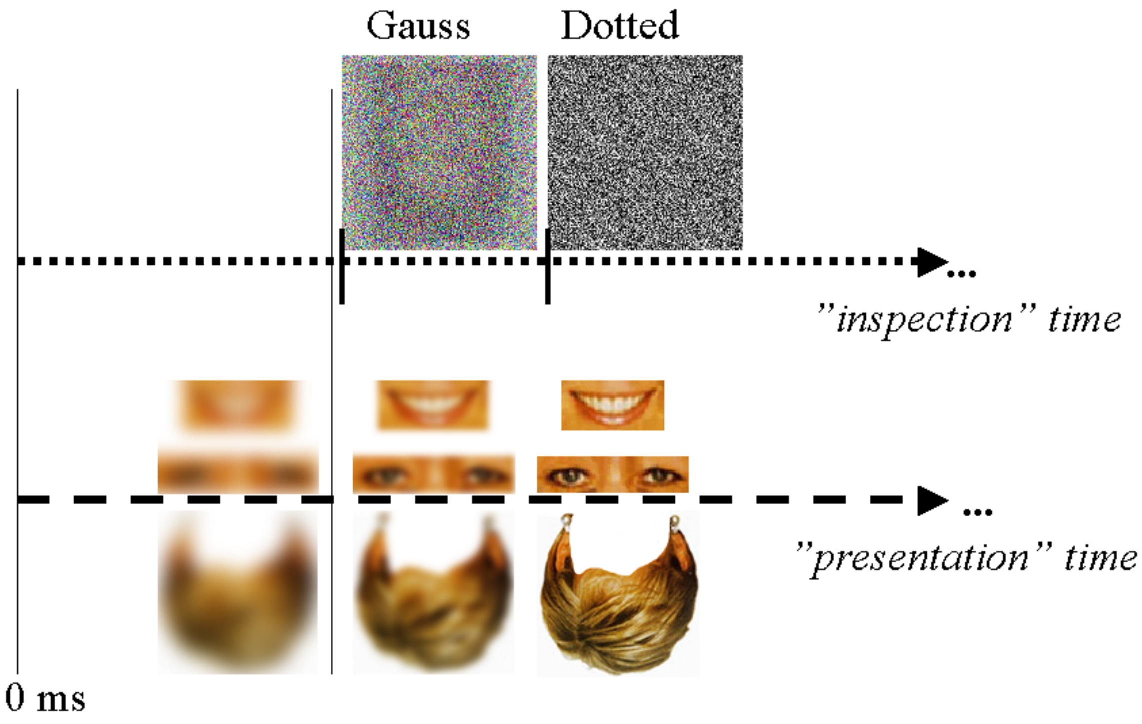
None of the factors were significant ($Fs_{1,29}<2.15$; $Ps>.1530$, n.s.). To our great surprise, the TAE vanished for this new experimental setting (i.e., the changed visual mask), demonstrated by a non-significant interaction between CLASS * PT ($F_{1,29}=1.31$; $P=.2623$, n.s.). The difference between Thatcher faces and original faces under the short PT (see Table 8-45 in the appendix) was reduced from Experiment 3a to Experiment 4 from a significant difference of 67.7 ms to a non-significant difference of only 3.9 ms. The latter difference was found to be not significant by a one-tailed *T*-test ($T_{29}=.097$, $P=.4618$, n.s.). At first sight, these results seem to indicate that the TAE is *not* a stable effect, because the nature of a visual mask has the power to make it vanish. Thatcher faces seem not to be recognized advantageously in terms of RT in comparison to original faces. Nevertheless, a deeper look at the data reveals a possible explanation. Figure 4-49 clearly shows that there is indeed no special effect for Thatcher faces under the short PT, but more so under the longer PT.

Figure 4-49: RT data (*yes-same*, ±3 *SDs*) for the *FULL* condition.



The TAE for PT=200 ms was also *not* significant (one-tailed *T*-test with T_{29}=1.27, *P*=.1067), but was now more substantial with a 53.0 ms difference. Thus, Hyp.1 was indeed falsified. By using an alternative Gauss mask, the Thatcher advantage effect (TAE) vanished! However, it seems that the alternative mask did not *extinguish* the TAE, but *transferred* the advantage effect to a later period of time. Presumably, the composed Gauss mask was very effective in masking. Effectiveness is defined here in terms of the capability to stop the further iconic processing of a visual stimulus. Thus, a Gauss filtered compilation of faces seems to be an ideal tool for deleting the iconic memory holding a facial stimulus, because it contains 'everything, but nothing in particular of a face'. In contrast, a random dotted pattern seems to be less effective. However, what does 'less effective' mean in terms of TAE? The following Figure 4-50 demonstrates the hypothesized influence of visual masking on the recognition process of a face.

Figure 4-50: Model for the relationship between psychologically relevant *inspection time* and physically relevant *presentation time*, using different qualities of masks.



This model illustrates the condition of psychologically relevant and physically given exposure times in the following way. Besides a physical presentation time (PT), there is a psychologically relevant ‘inspection time’, which is relevant in regard to the construct of the measure *IT* (Vickers et al., 1972). The nature and efficiency of the visual mask used for backward masking fix the inspection time. An ideal mask which totally covers the preceding stimulus without interfering it, would lead to the relation IT=PT. Under normal circumstances, this relation is never reached, therefore psychological IT is always longer than the physical PT, because of a persisting iconic memory (Coltheart, 1980). According to this logic, the weaker the masking effect, the longer may be the IT relative to the PT. In the illustration above, a less effective random dot pattern extends the IT more than a Gauss compilation mask does. Indications for the high efficiency of the Gaussian pattern were found in the post-experimental interviews and by my own perception: the face stimulus was deleted by the mask more effectively—the face items seemed to be ‘swallowed’ from the mask. Moreover, especially colored and textured information seems to be better masked with such a (colored) compilation Gauss-mask as described above. If it is the case that the Gauss mask used in Experiment 4 is really more ‘effective’, then it is plausible that the TAE is shifted to later time periods. This was indicated by an increase of TAE from PT=26 ms to PT=200 ms. Further studies will have to illuminate the specific role of the visual mask (see Breitmeyer, 1984).

Hyp.2–Enhanced TAE

For testing Hyp.2, the RT data for *yes-same IN*-faces had to be analyzed. A pre-check whether the RT data are analyzable turned out to be negative, because the Hit-rate (saying ‘yes’ to same trials) of Thatcher faces under the short PT condition was too low. A Hit-rate of only .476 indicates that the performance for this condition was *not* above the chance level of 50% (but also not below: two-tailed T_{29}=.67, P=.5097, n.s.). The task to detect faces from only inner parts being available was too difficult within a PT of 26 ms. Therefore, it seems to not be a fair test to further analyze the RT data for *yes-same* trials. Thus, Hyp.2 was not testable.

Hyp.3–Global precedence

The key assumption of Hyp.3 is that the identification of global structures like the outline of a face is available more quickly than local structures. Therefore, *OUT*-faces should be recognized faster than *FULL*-faces, and *FULL*-faces faster than *IN*-faces (for original faces). This was tested by a one-way ANOVA with TASK (*FULL*, *IN*, *OUT*) as between factor (see details in Table 8-46 in the appendix) and RT[62] (for *yes-same* trials) as dependent variable.
A significant effect of TASK ($F_{2,58}$=20.41; P<.0001) was revealed. All stages were found to be significantly different from each other (Scheffé post-hoc with Ps<.0230; more details are found in Table 8-47 in the appendix). *OUT*-faces were being recognized most quickly, which lends strong support to Hyp.3. Besides an assumed *precedence* of global structures, these results also indicate that not the *simplicity* in itself results in a fast response. If it were the case that a more simple or fluent structure causes fast RTs, not only *OUT*-faces should be recognizable faster, but also *IN*-faces, both compared to the *FULL*-condition, because *OUT*-faces as well as *IN*-faces contain less information and physically relevant space than *FULL*-faces. However, this proved not to be the case. The latencies for recognizing *IN*-faces were the longest of all TASKs (see Table 4-28). It seems to be very difficult to recognize faces within a very short PT if the outer structures are absent. The reason for this might be either the missing reference system or the artificial kind of presentation. This, again, is in accord with evidence found in Experiment 3a in regard to global recognition.

---

[62] As for testing Hyp.2, the calculation of RT data for *yes-same* trials in the *IN*-condition is rather problematic, because the recognition rate for this TASK was very low.

Hyp.4–Rotation

Hyp.4 predicted a reduced TAE for 45°-rotated *R45*-faces. As for the investigated Hyp.1, the specific visual mask used in this experiment seems to be more effective than the mask used in Experiment 3a. A possible reduction of TAE was analyzed for the long PT, because the TAE seems to be shifted to longer PTs due to the higher effectiveness of masking (see details in section about testing Hyp.1). Therefore, RT data (*yes-same* trials) for the *long* PT have to be compared between *FULL*- and *R45*-condition. The size of TAE in terms of RT was compared for both TASKS by a one-tailed *T*-test, which revealed a trend for the predicted reduction of TAE (*Diff*=60.64, T_{29}=1.61, *P*=.0593, n.s.). This is in accord with Hyp.4.

Summarizing these results, the specific usage of a composed Gauss mask in this experimental procedure left Hyp.2 untestable due to low recognition performance. Therefore, an additional study with different methodical parameters was conducted in Experiment 5.

4.4.4 Experiment 5: Inverted Thatcher faces with a dotted mask

Introduction.

This experiment aimed to test the *same hypotheses* as Experiment 4, but with an alternative experimental procedure. Detailed explanations of these hypotheses are given in the introduction and the overview of hypotheses in Experiment 4 (see page 134 et seqq.). Compared to Experiment 4, two specific parameters were changed.

On the one hand, Hyp.2 of Experiment 4 remained untestable due to low recognition rates in the *IN*-condition. It has already been discussed above that the used visual mask seemed to be very effective in deleting the iconic memory. Therefore, the random dotted mask used in Experiment 3a and Experiment 3b was used again.

On the other hand, the interrogative form used in the preceding Thatcher face experiments seems to be rather problematic. It might not be very intuitive to instruct participants to react to original faces and Thatcher faces in a different way, although they belong to the same face basis. This might lead the participant to an artificial responding strategy. Within this strategy, participants might not try to *recognize faces*, but might focus on *facial manipulations*. In order to remedy the weakness of this task, an alternative interrogative form was used in Experiment 5. Subjects were now asked whether the name and the succeeding face were compatible (i.e., the same), while any detected oddness was to be ignored. Such an interrogation seems to be more intuitive for an experiment testing face recognition phenomena.

Method.

Participants. As in all other Thatcher face experiments, thirty students participated in this experiment. The participants were undergraduate students (25 women) of the Freie Universität Berlin, who were given a course credit to fulfill course requirements. The mean age was 25.6 years (from 19 to 45 years). All participants had normal vision abilities or were corrected to normal vision and had not taken part in the former Thatcher face experiments.

Material. The stimulus material was the same as in Experiment 4. The only difference in material concerned the visual backward mask. Due to the delayed *Thatcher advantage effect* (TAE) in Experiment 4 by using a Gaussian mask, the random dot mask used for all the other experiments was now used again (see Figure 8-6).

Procedure. The time course of the trials and the sequence of experimental blocks was the same as in Experiment 4. The only difference in the procedure between both experiments was a changed format of the question asking for the participant's response. In the Thatcher experiments above, the question always referred to the *original* face of someone. Here, this question was simplified to 'Is this <pre- and surname>?'. If the participants ever had the feeling of a distorted face (Thatcher face), they should disregard this perception in their answer. This change of questioning reflects a more natural kind of query. In everyday life, a face is a face of *anyone*, but there are no 'original' or 'distorted' variations of a face. Additionally, with a high acceptance of Thatcher faces as ‘normal’ faces indicated by high *A’* and high *yes-same* rates in Experiment 3a and 4, only very few Thatcher faces are expected to be recognized as distorted anyway.

Hypotheses.

The hypotheses for this experiment are in accord with those of Experiment 4. Thus, further details are described in the corresponding section of Experiment 4. All tests were conducted with RT data, whereas TAE is defined as above as the RT-advantage of recognizing Thatcher faces against original faces.

Hyp.1–*Replication of Thatcher advantage effect (TAE):* The TAE (for *FULL*-faces) should be replicated even with a changed and more natural kind of questioning, because it is a stable and reliable effect.

Hyp.2–*Enhanced TAE for IN-faces:* TAE(*IN*) > TAE(*FULL*) for short PT.

Hyp.3–*Global precedence:* RT(*OUT*) < RT(*FULL*) < RT(*IN*).

Hyp.4–*Rotation:* TAE(*R45*) < TAE(*FULL*).

Results & Discussion.

Again, only familiar faces (93.7% of all faces) were included in the subsequent analyses. The overall RT data of all TASKs are shown in Table 4-29. As in Experiment 4, the RT data were limited to ±3 *SD*s around the individual RT mean of every subject.

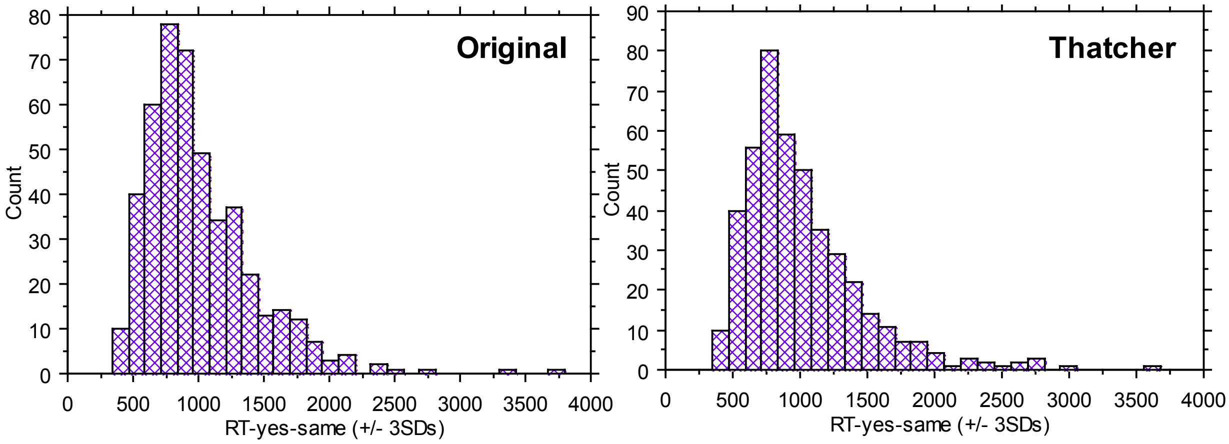
Table 4-29: RT-data for *yes-same* trials (in ms). *SD*s are given in square brackets. RTs were restricted by the criterion of ±3 *SD*s.

TASK	*PT*	*Original*	*SD*	*Thatcher*	*SD*	*TAE*
FULL	short	1061.7	[265.1]	1006.0	[275.3]	55.7
	long	971.7	[238.0]	1002.6	[255.3]	-30.9
IN	short	1154.2	[405.8]	1085.2	[290.8]	69.0
	long	1101.8	[310.0]	1167.5	[343.6]	-65.7
R45	short	914.1	[325.1]	880.8	[225.3]	33.3
	long	803.5	[186.6]	837.2	[274.8]	-33.7

OUT	*PT*	*OUT-faces*	*SD*
	short	987.6	[319.2]
	long	868.2	[228.9]

As a first exploration of data, the distribution of RT data was analyzed in order to exclude any artifacts in the RT pattern due to a possible inhibition effect. An inhibition effect might surface if people have to answer ‘yes this is the target person’ while looking at a very similar but *distorted* (thatcherised) face. It is possible that the detection of a Thatcher face leads to an inhibition of response for those faces. If such an inhibition effect occurred, then the RT distribution (for *yes-same* trials, ±3 *SD*s) would not follow a typical left-skewed mono-modal one, but would instead reveal a bimodal distribution. A look at the RT data of the Original and the critical Thatcher faces revealed no particularly irregularities (see Figure 4-51).

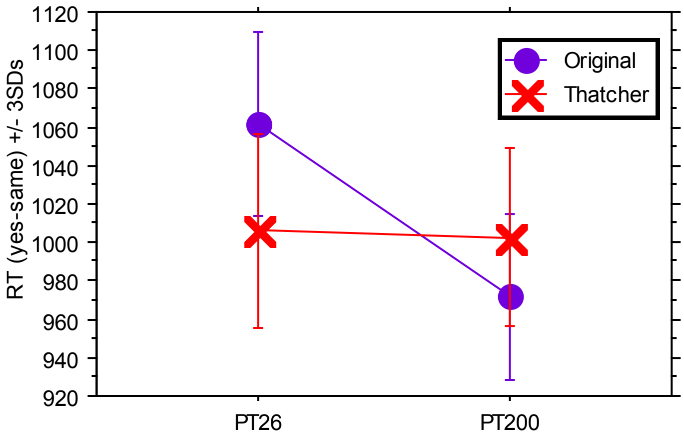
Figure 4-51: RT distribution for *yes-same* trials (±3 *SD*s). On the left, the original face condition is seen, on the right the Thatcher face condition is presented.



Hyp.1–Replication of Thatcher advantage effect (TAE)

It was tested whether the TAE, demonstrated in Experiment 3a, also occurred when using a different question. A repeated measurement two-way ANOVA with the within-subjects factors CLASS (Original vs. Thatcher) and PT (short: 26 ms vs. long: 200 ms) and with RT data (*yes-same* trials, ±3 *SD*s) of *FULL*-faces as dependent variable was conducted (see Table 8-48 in the appendix). The interaction between CLASS and PT was only revealed as a trend ($F_{1,29}$=3.22; P=.0830, n.s.). As shown in Figure 4-52, the plot of the RT data is very similar to that of Experiment 3a, but now the interaction fails to be significant, whereas in Experiment 3a this interaction had been significant.

Figure 4-52: RT data (*yes-same* trials, ±3 *SD*s) split by CLASS and PT. Only *FULL*-faces were included.



A deeper analysis of the hypothesized TAE also revealed a trend for a difference between original condition and Thatcher condition under the short PT for *FULL*-faces (one tailed T_{29}=1.61, P=.0595, n.s). This trend was rather large with a TAE of 55.7 ms (see full data in Table 4-29).

Thus far, only *yes-same* trials were analyzed, because in the former experimental designs the definition of a ‘correct’ response was problematic due to the interrogative form. With the question used in the present experiment, this was no longer the case. Thus, in addition, the RTs for *correct trials* were analyzed. Before calculating RTs for correct trials (of *FULL*-faces), the distribution of RT data was examined again. Once more, no deviation from an assumed mono-modal distribution was found (see Figure 8-13 in the appendix). A two-way repeated measurement ANOVA with CLASS and PT as within-subjects factors (see details in Table 8-49 in the appendix) showed a similar trend for interaction between CLASS and PT ($F_{1,29}$=3.39; *P*=.0760, n.s.). Besides this interaction, the TAE was again marginally significant (one tailed T_{29}=1.63, *P*=.0569, n.s) with a size of 41.1 ms (see full data in Table 8-50 in the appendix).

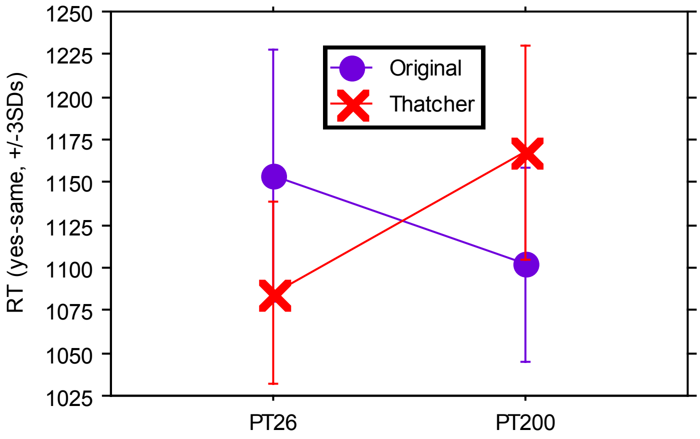
This combined evidence demonstrates that the TAE is reduced by an alternative form of questioning, but there is still a very strong trend for it. Therefore, the TAE can be assumed to be relatively stable, as predicted by Hyp.1.

Hyp.2–Enhanced TAE for IN-faces

In order to test Hyp.2, the RT data (for *yes-same* trials, ±3 *SD*s) for *IN*-faces were investigated (see full data in Table 4-29). Therefore, a repeated measurement ANOVA with CLASS and PT as within factors was conducted for RT data of *IN*-faces. Again, the interaction of CLASS and PT was marginally significant ($F_{1,29}$=4.02; *P*=.0543, n.s.). Although the TAE for *IN*-faces was numerically much larger than for *FULL*-faces (69.0 ms vs. 55.7 ms; see details in Table 4-29), the effect was not statistically significant due to the higher variances (one tailed *T*-test with T_{29}=1.18, *P*=.1242, n.s).

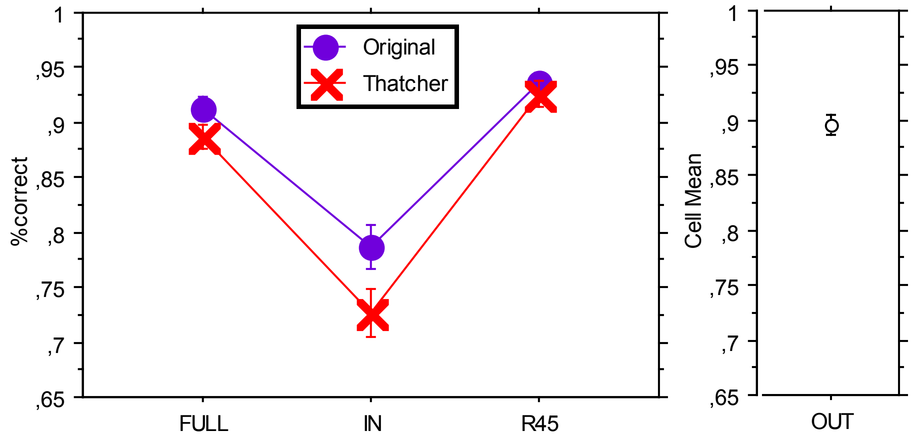
However, as illustrated in Figure 4-53, the marginally significant interaction follows the prediction with a (positive) TAE under a short PT and a negative TAE under the longer PT condition.

Figure 4-53: RT data for *IN*-faces (*yes-same* trials, ±3 *SD*s).



Thus, there were only numerical indications in support for hypothesis Hyp.2, which could not be corroborated statistically. However, what seems obvious is that once again *IN*-faces were very hard to recognize within 26 ms. The rate of correctly identified original *IN*-faces was only .787, which was the lowest rate of all TASKs (see Figure 4-54).

Figure 4-54: Correct identification rates for the different TASKs. On the left side, all the conditions that could be split by CLASS. Due to the lack of inner features, the CLASS condition was not available for *OUT*-faces. For comparative purposes, however, the *OUT*-condition data is also given (on the right). The referring data for this illustration will be found in Table 8-52.

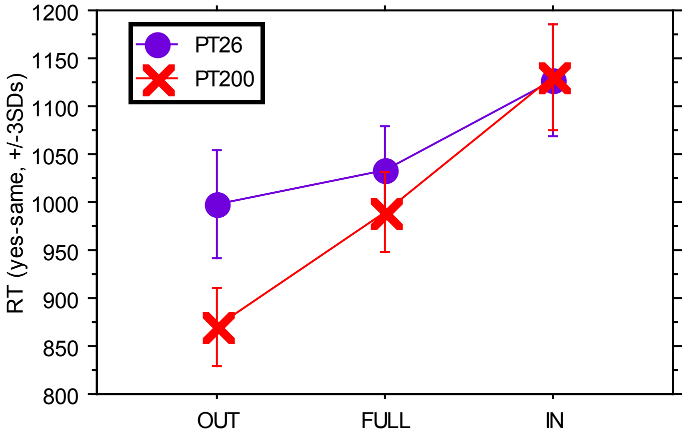


The relatively weak recognition performance for *IN*-faces was further investigated by a one-way repeated measurement ANOVA (see Table 8-53 in the appendix) with the factor TASK (*FULL, IN*, *OUT, R45*). The factor TASK was significant ($F_{3,87}$=38.6; P<.0001). A post-hoc Scheffé test revealed the *IN*-condition to be weaker than any other condition (see Table 8-54 in the appendix). The weak performance reduced the number of calculable incidences of *IN*-data. Additionally, the specific difficulty of the *IN*-task heightened the false alarm rate. Both difficulties combined increased the noise of the data and therefore lowered the statistical power of the test. This is the reason that the *IN*-condition remains problematic.

Hyp.3–Global precedence

For testing Hyp.3, a repeated measurement ANOVA (see Table 8-56 in the appendix) with the two within factors TASK (*FULL*, *IN*, *OUT*) and PT for the RT data (yes same, ±3 *SD*s) was conducted. There was a significant TASK factor ($F_{2,58}$=15.51; P<.0001) with latencies of the *IN*-condition being longer than those of the *FULL* and *OUT*-condition (see Scheffé post-hoc tests in Table 8-57 in the appendix). This is partly in accord with the prediction of Hyp.3. Nevertheless, the difference between *OUT*-faces and *FULL*-faces failed to reach significance. Interestingly, the lack of difference was only caused by RT data for the short PT. Figure 4-55 shows the RT data split by PT. It is obvious that the RT-predictions of Hyp.3 were confirmed for the processing of faces in the PT-long condition. For the long PT, there was a significant order from *OUT*-faces being recognized most quickly. Next came *FULL*-faces, while *IN*-faces revealed the longest latencies (see split up Scheffé post tests in Table 8-58 in the appendix).

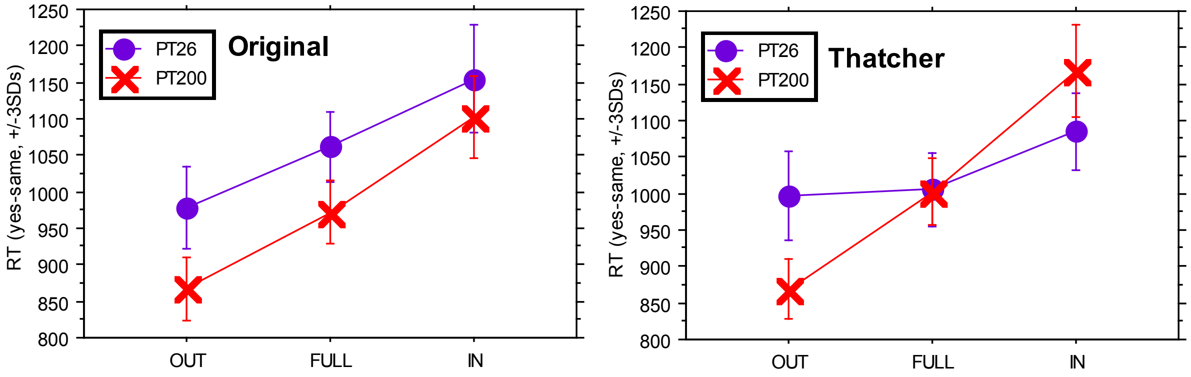
Figure 4-55: RT data (*yes-same* trials, ±3 *SD*s) split by PT for testing Hyp.3 (see full data in Table 8-55 in the appendix).



What might have caused the dissociation between the two different PTs indicated by a interaction of TASK and PT ($F_{2,58}$=3.73; P=.0300)?

A further analysis revealed that this interaction effect was based on a dissociation for Thatcher faces, but not for original faces (see Figure 4-56). The recognition speed of very briefly presented Thatcher faces did not differ between the single tasks (*OUT*, *FULL*, *IN*), but showed the predicted order when the presentation was extended to 200 ms (revealed by Scheffé post-hoc tests in Table 8-59 in the appendix).

Figure 4-56: RT data (yes same, ±3 *SD*s) split by CLASS and PT for testing Hyp.3. There was no genuine Thatcher/Original condition for *OUT*-faces due to the omitted inner facial parts, but for comparing the data among the different conditions, a random selected half of the *OUT*-data was taken for 'Original faces' and the random selected other half for 'Thatcher faces'.



These results confirm the findings of the TAE discussed above. It was found once again that a briefly presented Thatcher face possessed an advantage in the recognition of a face due to the already upright local features, but this advantage was lost when the PT increased. As shown in the left plot of Figure 4-56, there is a global precedence in terms of RT for original faces. For Thatcher faces, the TAE under a short PT seems to be so advantageous that Thatcher faces are even recognized as quickly as simple outlines.

The same argumentation seems to be valid for thatcherised *IN*-faces versus thatcherised *OUT*-faces. When faces had been presented very briefly, TAE was most effective and therefore even compensated the global precedence.

Nevertheless, the high accuracy rate and the high speed of recognizing *OUT*-faces demonstrate that global structures are important for face recognition. This is particularly the case when a face is presented under difficult conditions such as very short PTs.

However, it seems that global structures are not the only and not the *most* important information recognized within the first milliseconds of face processing. Moreover, the demonstration of possible compensation of global precedence by local feature analyses in Thatcher faces reveals that local feature identification processes might be occurring in parallel with global processing. This lends support to Hyp.3.

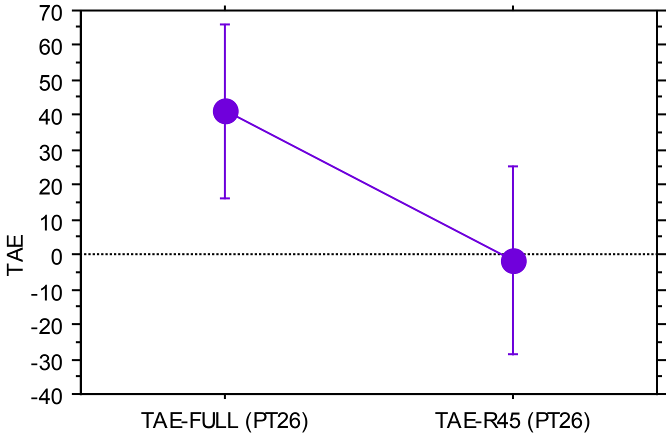
Hyp.4–Rotation

In order to test Hyp.4, the RT data for *R45*-faces had to be analyzed.

As mentioned before, the TAE was explained by a local feature analysis advantage at brief presentation. For inverted Thatcher faces, the local components eyes and mouth do not have to be rotated mentally, because they are already in an upright orientation. Therefore, the TAE is predicted by Hyp.4 to be reduced when faces are deviated by 45° from full inversion in condition *R45* (see illustration of the *R45*-condition in Figure 4-47).

To prove this hypothesis, the TAE for PT=26 ms was calculated for *FULL* as well as for *R45* faces by subtracting the RT data for *correct trials* of original faces from thatcherised faces as already described in Experiment 3a (see RT data for correct trials in Table 8-60 in the appendix). Although the TAE was indeed numerically much greater for the *FULL*-condition than for the *R45*-condition (41.1 ms vs. -1.6 ms; see Figure 4-57), this difference turned out not to be significant (one-tailed *T*-test with T_{29}=1.10, *P*=.1409, n.s.) due to high variances.

Figure 4-57: *Thatcher advantage effect* (TAE for *correct trials* under PT=26 ms) for *FULL*-faces vs. *R45*-faces (see full data calculation in Table 8-60 in the appendix).



However, a trend for the difference of the TAE between *FULL* and *R45* was in accord with Hyp.4. This trend does not constitute direct evidence for the early local feature analysis hypothesis. However, the decrease of TAE for *R45*-faces in combination with the RT data discussed above indicates that local features are processed advantageously, yet in a very early processing stage.

Additionally, these findings are in accord with the hypothesis that internal features are very important for the recognition of faces. This is also in line with early work of Vicki Bruce, who found that repeated interactions with a person's face leads to a structural code, which emphasizes the internal, communicative features—especially the eyes and the mouth (Bruce, 1982). This hypothesis seems very reasonable, because the outer part, especially the hair (styling), can change very rapidly or can be easily covered by hats or other clothes.

Nevertheless, a very simple 2AFC[63] or a simple Yes-No-decision (as used here) is already possible, based solely on the recognition of the outer parts, revealed by fast and accurate processing of *OUT*-faces. It is questionable whether such an RT pattern would also be found for more complex recognition tasks, because the presentation of outer structures *only* do not seem to provide valid cues for a person's identity, particularly for higher cognitive recognition tasks.

To summarize the findings of the present experiment, two things have to be mentioned. On the one hand, global structures seem to be important and effective for simple face recognition tasks. On the other hand, local feature analysis processes seem to occur simultaneously with global identification processes. During brief presentation, the processing of the eyes and mouth region might be very helpful for face recognition, probably due to their high physical information load.

4.4.5 Summary of experiments with familiar faces

The experiments presented in the previous sections have investigated the early processing of local facial features and of global face structures.

In order to be able to dissociate local and global processing, so-called *Thatcher faces* have been used as stimuli. All stimuli were based on frontal photographs of famous faces, which were all familiar to the participants. In Thatcher faces, the eyes- and mouth-regions are turned upside-down. This kind of manipulation is recognizable immediately. By inverting the whole Thatcher *Gestalt*, the alterations vanish and the Thatcher faces look like normal faces. This characteristic was utilized to dissociate local and global processing.

Experiment 3a revealed that shortly presented (inverted) Thatcher faces (26 ms) were recognized faster than (inverted) normal faces ('original faces'). This RT effect was called 'TAE'—*Thatcher advantage effect*. It was assumed that beneficial local feature analyses were responsible for this RT advantage. Additionally, it was found that holistic or template-like face processes did not come into play until later.

Experiment 3b, the only experiment operating with upright stimuli in this series, demonstrated that the processing of upright faces is much more efficient than that of inverted faces. Thatcher-alterations were already detectable after a presentation of only 26 ms.

The next two experiments, Experiment 4 and Experiment 5, further investigated the results of early local feature analyses found in Experiment 3a. The TAE was found to be stable in Experiment 5, although the interrogative form of the task had been changed. In Experiment 4, a trend for a TAE was shifted to the longer PT by using a different visual mask.

Moreover, a global precedence in processing faces was discovered. The recognition of global structures like the outline of a face was very fast and highly accurate (cf. Sergent, 1984b). It was concluded that simple global structures might be very efficient for basic recognition tasks. Nevertheless, local feature analyses could be similarly advantageous for the recognition process. Therefore, the assumption of a *general* global precedence was not supported.

Interestingly, the recognition of local structures *only*, tested by presenting only internal parts of faces, was surprisingly time consuming. It took a longer time to recognize the exclusive internal part of a face than to recognize the whole face, although an internal version contains less information than a full face. Thus, not the *quantity* of facial stimulus information is relevant for face processing, but the *quality* of the presented material.

---

[63] 2AFC: *two-alternative forced choice*. An experimental procedure in which the user is constrained to select one solution from two alternatives.