"If men had been born blind, philosophy would be more perfect, because it would lack many false assumptions that have been taken from the sense of sight"

(Galilei, 1983)[3]

2 Face Recognition

2.1 Why investigating face recognition?

Face recognition is a wide field, investigated by psychologists, computer specialists, biologists, anthropologists, neuro–scientists. Finally, yet importantly, face recognition is of some interest for the man of the street. The question whether a baby's face has again changed again and whom of the several relatives it now resembles the most, is a welcome subject for discussion on every family celebration.

Thus, this field of research receives particular attention. We are only able to explain very small parts of the recognition process, and thus only see the outcome of such a process, but not the sub-steps that lead to its completion. It is a fact that we are able to recognize thousands of faces and distinguish them properly. As 'face experts', we regard every face, at least of our own ethnic group, as having its own idiosyncratic look. A face can be divided into many facial dimensions where each dimension consists of *n* different values. Let us assume, for example, that a face only consists of ten natural dimensions (e.g., hair, forehead, eyes, eyebrows, nose, mouth, chin, skin, ears, form) with ten values per dimension. Such a simple assumption may create up to $10^{10} = 10$ billion possible faces (cf. Rakover & Cahlon, 1998), which is already more than the world's total human population. However, as we will see later, our face recognition system is much more sensitive than it could be described by 10 differentiated dimensions with 10 values. Because of man's extreme sensitivity to faces and the capability of encoding them, deeply, faces are the ideal tool for discriminating among different people in order to identify them (Lockhart & Craik, 1990).

In the past 20 years, there has been a dramatic increase in the research on face recognition. Many bibliographies about this field of interest have been compiled, for instance by R.J. Baron (cited by Bruce & Young, 1986), who has listed over 200 relevant papers in 1979. A renewed version of Zhao et al. (2000) also lists about 200 top relevant articles, and a search made by the psychological retrieval system *PsycInfo* found even 1169 papers on the topic of "Face recognition". The international citation-finding system *Citeseer* (citeseer.nj.nec.com)

[3] Originally published in 1610.

found 1249 relevant places (both retrievals carried out in September 2002). However, up to now there are relatively few papers dealing with the *processing* of face perception, and especially with the very early stages of this process. The *processing* of faces will be the main focus of this work.

2.2 Expertise

Face processing of adults is based on innate mechanisms, and is a product of years of experience. There is no doubt that face processing is a human skill in which most adults are real experts. In the following section, expertise in general and *face expertise* in particular will be investigated.

2.2.1 The nature of expertise

The term *jizz* is used by veteran birdwatchers to describe their flash of instant recognition of a bird based on its color, shape, and movement (Tanaka & Curran, 2001). A very similar phenomenon can be found in our everyday treatment of faces. We are by a practical definition of the words real 'face experts'. It seems that our perceptual system generally narrows with development (Gauthier & Nelson, 2001), and with this narrowing our expertise grows (Schwaninger, Carbon, & Leder, in press). *Prosopagnosia*, the clinical inability to recognize faces, is direct evidence that face recognizing is dissociable from other object processing capabilities. Mostly, prosopagnosians suffer from a brain injury; whereas they are still able to see objects and recognize parts of faces, they have lost the ability to recognize whole faces even of familiar people (Young & De Haan, 1995; Takamura, 1996).

Our expertise of face recognition is specialized to the class of human faces of our own race[4]. This specialization is also found in animals. Sheep, for instance, whilst being extensively exposed to interactions with humans, were unable to identify them with the same 'expert' methods that they used to discriminate members of their own kind (Peirce, Leigh, DaCosta, & Kendrick, 2001).

What is an expert?

Camerer and Johnson (1991) define an expert as "a person who is experienced at making predictions in a domain and has some professional or social credentials" (Camerer & Johnson, 1991, p.196). What seems to be essential is that experts show a highly selective search, which means that they do not search more than non-experts (De Groot, 1965; Gobet, 1998). It seems that experts "see the solution". With their immense knowledge, they can shorten their search strategies and thus optimize the desired solution. Moreover, experts consider putatively diagnostic subsets of variables which in reality may not be strongly related to the outcome (Camerer & Johnson, 1991). Experts seem to use less, rather than more, information than novices do! The reason for this fact is that they see the material at a higher level than novices; they are able to cut the material into larger meaningful units, often referred to as *chunks* (cf. Miller, 1956).

Experts can develop mechanisms that allow a rapid storage of information into long-term memory (LTM), which is the principle of retrieval structure (Ericsson & Kintsch, 1995; Chase & Ericsson, 1982). With a well-organized and efficient retrieval structure, consisting of a set of retrieval cues, a fast storage of information is possible, because *only values* need to be encoded, whereas the structure itself is already stored. A prominent example of an expert using such a retrieval structure is 'DD', a subject who was able to recall up to 106 digits dictated at the rate of one second each (Chase & Ericsson, 1982).

---

[4] *Race* will not be defined here as a biological taxonomy. Instead, it is defined within the tradition of social psychology in terms of the wider ethnic, cultural or social group (see Reber & Reber, 2002).

Experts' knowledge is partly coded as such meaningful entities; for example, it has been estimated that chess grandmasters have a repertoire of about 10,000 – 100,000 chunks (Simon & Gilmartin, 1973; Chase & Simon, 1973), which is not only larger than that of novices but also better organized (Chi, Feltovitch, & Glaser, 1981). On the other hand, poor recognizers tend to pick a single feature or small set of distinctive features (Biederman & Kalocsai, 1998).

Neuropsychological evidences

Neuropsychologists have found enhanced activity in face-selective extrastriate areas with the development of expertise for novel objects (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1995). It could be demonstrated that approximately 164 ms after presentation, objects of expertise can be differentiated neurologically from objects of lesser-known categories (Tanaka & Curran, 2001). Car and bird expertise also recruit face-selective areas (Gauthier, Skudlarski, Gore, & Anderson, 2000). At the spatial resolution of *functional Magnet Resonance Imaging* (fMRI), peaks of activity for faces, birds or cars in experts, could not be dissociated (Curby & Gauthier, subm.). Thus, the hypothesis that different expertise domains recruit the same underlying processes seems plausible. For example, activity for *Greebles*[5] in the fusiform 'face area' (FFA) (Kanwisher, McDermott, & Chun, 1997) increased with expertise (Gauthier & Tarr, 2002). In other words, the FFA may not be specific for faces *per se*, but rather only for the operations we perform typically and by default when perceiving faces (Gauthier, 2000; Tarr & Gauthier, 2000; Gauthier & Logothetis, 2000).
A necessary condition for becoming an expert is extensive training. The neurons needed for the fine discrimination processes have to be tuned for fast and accurate identification of individuals (Williams, Gauthier, & Tarr, 1998). Changes in the timing and morphology of the ERP responses lead to the conclusion that cortical specificity for face-processing increases across development (Mondloch, Le Grand, & Maurer, 2002; Taylor, Edmonds, McCarthy, & Allison, 2001; Bruce, Desimone, & Gross, 1981).
From an evolutionary standpoint, it makes sense that humans should have developed an area of the brain that is specific of the processing of faces or highly similar objects, to establish specialized mechanisms, which are superior to other tasks. The ability to recognize particular faces is crucial to many aspects of social interaction and necessary for survival, for example to distinguish between family and strangers, as well as between friends and enemies.

Face-selective areas

The remarkable speed of recognition in the primate brain allows for only very brief processing times at each stage in a pipeline, containing only few stages (Mel & Fiser, 2000). Therefore we need to have specialized brain areas, that can identify or pre-categorize faces fast and accurately as being highly important biological and social entities. The recognition of faces plays a particularly important role in the identification of faces as being those of enemies or friends. Indeed, statistically significant differences between (schematic) face and non-face stimuli already appeared around 135 ms from stimulus onset (Yamamoto & Kashikura, 1999), which is an indicator for a very early pre-categorization of faces against other objects (cf. Seeck & Grüsser, 1992). Such so-called *receptive fields* are sensitive to some localized spatial configurations of image cues but invariant to one or more spatial transformations of its preferred stimulus (Mel & Fiser, 2000). Logothetis, Pauls and Poggio (1995) assumed that receptive fields are not hard-wired from the moment of birth on, but are 'tunable' to relevant stimulus classes. They conducted single-cell experiments in monkeys which suggest that the selectivity of neurons can be tuned by experience with artificial objects similar to face cell selectivity (cf. Gauthier & Tarr, 1997; Gauthier, Williams, Tarr, & Tanaka, 1997). They extensively trained monkeys to identify novel 3D objects (~600 000 trials for a given object category) and

[5] Artifical stimulus class, which is highly complex. Often used by the Gauthier research group (e.g., Gauthier & Tarr, 1997) for testing putative face specific effects.

found a population of *inferotemporal* (IT) neurons that responded selectively to novel views of previously unfamiliar objects. Importantly, the tuning properties of cells to novel objects can hardly be explained by an innate predisposition and are more consistent with the type of plasticity postulated by the expertise hypothesis (Logothetis et al., 1995). Other data showed that activations associated with animate objects or living things (i.e. animals, faces) cluster in the more lateral aspect of the fusiform gyrus, whereas activations associated with inanimate or man-made objects (i.e. tools, houses) cluster in the more medial aspect of the fusiform gyrus (Martin & Chao, 2001). Seeck and Grüsser (1992) also found some specific processing of 'face data'. Face-related components are contained in the early sequence of a negative peak (80-110 ms), followed by a large positive peak at about 150-200 ms and a rapid consecutive negativity. So it can be concluded that neural structures apparently specialized for face analysis have been identified with electrophysiological recordings, as well as with neuroimaging, in both monkeys and humans (see Leopold, O'Toole, Vetter, & Blanz, 2001).

The N170 occipito-temporal component

In the neuropsychology of face recognition, the *N170* occipito-temporal component plays a very important role. It can be reliably demonstrated by scalp electrophysiological (ERPs) studies that face and object processing differ approximately 170 ms following stimulus presentation with more negativity in the right hemisphere than in the left hemisphere (Tanaka & Curran, 2001), which is why this component is briefly called 'N170' (Rossion, Gauthier, Goffaux, Tarr, & Crommelinck, 2002).

The N170 reflects an early stage of face processing and is delayed and enhanced for inverted faces (Bentin, Allison, Puce, Perez, & McCarthy, 1996). Besides, the N170 was strongly attenuated for cheek and back views of faces relative to front and profile views, demonstrating that it is not merely triggered by 'head detection'. The same attenuated and delayed N170 components resulted for faces lacking internal features as well as for faces without external features, suggesting that it is not exclusively sensitive to salient internal features. Precisely speaking, the N170 *amplitude* was not affected by the presence or absence of eyes, but N170 was *delayed* in response to faces without eyes (Eimer, 1998). It is suggested that the N170 is linked to stages of structural encoding, where representations of global face configurations are generated in order to be utilized by subsequent face recognition processes (Eimer, 2000).

Based on the fact that patients with prosopagnosia either fail to demonstrate an enhanced N170 to faces (Eimer & McCarthy, 1999) or demonstrate a nonselective enhanced N170 to both face and nonface stimuli (Bentin, Deouell, & Soroker, 1999), N170 was used to be seen as direct indicator for *face* processing. Newer data criticized this view by demonstrating typical 'face specific' N170 activity also with non-face material (Rossion et al., in press).

Criticism of selective neurological patterns

Rossion and Gauthier (2002) could also demonstrate that the *topographical* distribution of the N170 is highly similar for faces and other object categories (Rossion & Gauthier, 2002, Figure 2). In any case, given the poor spatial resolution of ERPs, it is impossible to know whether ERP components in two conditions, even if identical in topographical distribution, latency and amplitude, represent the activity of exactly the same neural network (Rossion, Curran, & Gauthier, 2002).

Therefore, Rossion et al. (2000) hypothesized that the larger amplitude N170 for faces as compared to objects does not provide compelling evidence for early face-specific processes for three reasons:

1. The N170 may be due to low-level visual differences between faces and objects (e.g., spatial frequency);
2. The N170 amplitude can be larger between two non-face object categories (e.g., chairs and cars) than between faces and other object classes;

3. Face inversion and prosopagnosia seems to only minimally influence N170 activity (cf. Rossion & Gauthier, 2002). Indeed, in these cases the N170 component arises robustly and sometimes larger than for ordinary face recognition (cf. Rossion, Gauthier et al., 2002). These results appear at odds with the strong effect of inversion on face recognition performance.

Contrary to this view, the N170 to faces is immune to strategic manipulations, whereas the N170 to other categories can be influenced by task or attentional factors, suggesting that it does not represent the same process (Carmel & Bentin, 2002).

2.2.2 It is a long and winding road to become a face expert

It takes a long time to become an expert. For instance in the field of chess it takes about ten years to world-class level (Simon & Chase, 1973). The same is commonly said about face expertise, whether one is a child or an adult (Carey & Diamond, 1994). Only after a developmental phase of no less than 10-14 years with an ongoing cortical restructuring (Bartlett & Searcy, 1993) is the full expertise ready to work properly (Le Grand, Mondloch, Maurer, & Brent, 2001; Mondloch et al., 2002). Marked improvement in recognition tasks is observed between the ages of 2 and 10 (Carey & Diamond, 1977). Therefore, it is justified to speak of a downright *face recognition ontogeny*[6].

This brain development can be interpreted as a training of specialized brain areas or can be seen as a selection mechanism based on an unspecific 'resonance' between the sensorial input and the brain (Pallbo, 1999).

Face Expertise is prepared but has to be developed

On the one hand, numerous studies indicate that newborns will preferentially turn their eyes or head towards face-like stimuli within minutes or hours after birth (Goren, Sarty, & Wu, 1975; Johnson & Morton, 1991; Morton, Johnson, & Maurer, 1990; Mondloch et al., 1999; Simion, Valenza, Umiltà, & Dalla Barba, 1998; Valenza, Simion, Cassia, & Umiltà, 1996). Nevertheless, it does not seem clear whether this is true only of dynamic or even static face stimuli (Morton & Johnson, 1991). Two hypotheses have been offered to explain this early preference: the *structural* hypothesis and the *sensory* hypothesis. The structural hypothesis—sometimes called *conspec* hypothesis (Johnson & Morton, 1991; Morton & Johnson, 1991)—maintains that facelike patterns are special because of the structural organization of their internal features. The sensory hypothesis—sometimes called *conlearn* hypothesis (Kleiner & Banks, 1987)—assumes that, in principle, facelike patterns are not special but, as other classes of stimuli, are preferred as a result of the general properties of the early stage of visual processing.

By using the habituation-dishabituation paradigm (e.g., Johnson, 2001), it was possible to demonstrate that babies' capability of processing faces is not limited to face recognition based on similarity in pattern arrangement, only, but is based on a representational level (Roder, Bates, Crowell, Schilling, & Bushnell, 1992). Therefore, many researchers argue that there must be a genetically hard-wired brain mechanism specific to face processing (Johnson & Morton, 1991; Mondloch et al., 1999; Simion et al., 1998). Indeed, specific brain regions that respond preferentially to faces have been found in adult humans and in infant monkeys (Kanwisher et al., 1997; Desimone, 1991; Rodman, 1994).

On the other hand, numerous studies also demonstrate rapid experience-driven development of face processing abilities, even in the first few hours and days after birth (Walton, Armstrong, & Bower, 1997; Walton & Bower, 1993; Bushnell, Sai, & Mullin, 1989; Field, Cohen,

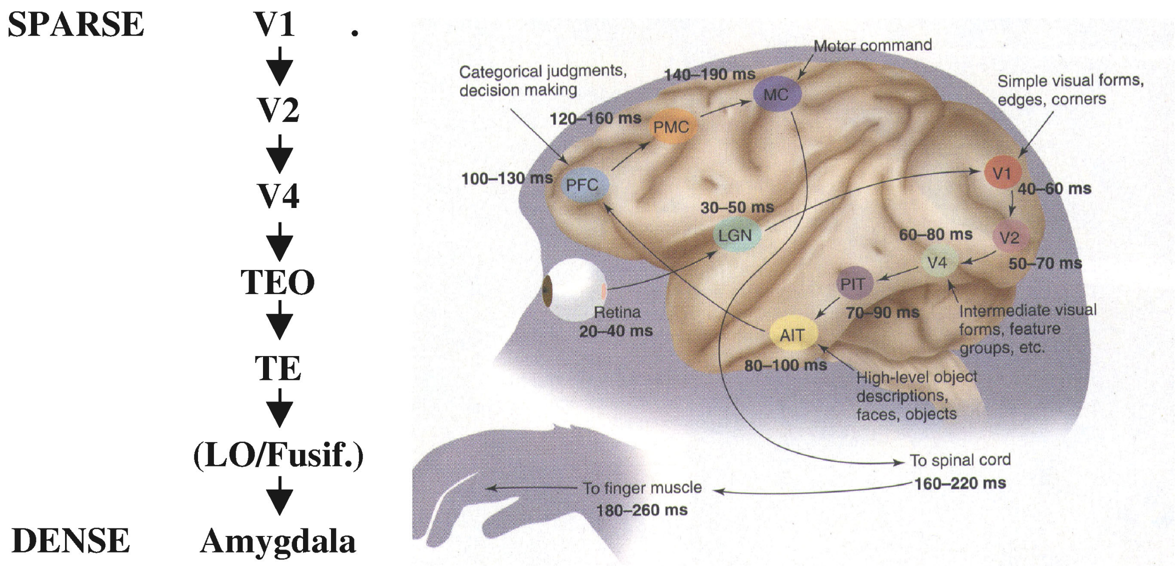
[6] The study of the ontogeny of face recognition has a long and illustrious history, dating at least back to Charles Darwin (1872).

Garcia, & Greenberg, 1984), which are paralleled by a massive increase in the volume of the human brain (Johnson, 2001). Such plasticity appears to be the rule, rather than the exception, in both the developing and the adult nervous system (Buonomano & Merzenich, 1998; Bednar & Miikkulainen, 2001; Bednar & Miikkulainen, 2000).

2.3 Impressive abilities

No video system or computerized camera, no matter how sophisticated, can match the impressive ability of the human visual system comprehend an infinite variety of images. This capability is due to the brain's capacity to process huge amounts of information simultaneously (Livingstone, 1988) and to extract the relevant information from the redundant or irrelevant at a very early stage (Marr, 1982). In the first 150 ms of image processing, the visual information goes through at least 10 different areas (Lewis et al., 1981; Thorpe & Fabre-Thorpe, 2001) in a rapid feed-forward *pipeline* architecture (Perrett, Rolls, & Caan, 1982; Perrett, Mistlin, & Chitty, 1987; Thorpe & Fabre-Thorpe, 2001; Delorme, Richard, & Fabre-Thorpe, 1999). In Figure 2-1 a plausible route between the retina and the muscles of the hand during a categorization task is depicted.

Figure 2-1: A plausible route between the retina and the muscles assumed by Thorpe and Fabre-Thorpe (2001) for a categorization task, confirmed by many empirical evidences. It is obvious that a lot of information is processed in the first 200 ms and thus many higher cognitive systems are already involved.



From Thorpe & Fabre-Thorpe, 2001

The human face enjoys an important and unique status in psychological research. As a *par excellence* prototype of complex multidimensional visual patterns, which everyone is familiar with, it has become a well-suited stimulus for the study of visual perception and the processes underlying the combination of component parts into a meaningful whole. Contrary to the research topics of the 70s, the main material used to understand processing hypotheses, is therefore not restricted anymore to the investigation of words or schematic objects, but has been expanded to much more complex facial stimuli. As an essential medium of interpersonal relationships, a human face conveys and reveals a wide variety of information about an individual, and the extraction and interpretation of this information require elaborate and refined perceptual skills that few other categories of objects call for. These skills have attained a very high level of proficiency, as evidenced by the capacity to identify a person in less than a second (Sergent, 1989) despite the close similarity among faces and the necessity to process several features, the arrangement of which uniquely defines a particular face. While the operations to interpret a face appear to be performed automatically and very fast, the brain must nonetheless solve a series of difficulties inherent in the perception. Some of these difficulties may be worth pointing out. For instance, the ever-changing appearance of a face, which can take different expressions and is to be perceived from different viewpoints. Moreover, the non-homogeneity of its surface that results in different patterns of brightness with a change of illumination. Furthermore, the great similarity between faces, and therefore the necessity to detect small variations between them, as well as within them.

While the most obvious functions of the face recognition system are to *detect* a face (e.g., Lakshminarayanan, Bhatia, Samal, & Welland, 1997), *recognize* it (e.g., O'Toole, Bülthoff, & Walker, 1995) and most importantly to bear a person's *identity* (e.g., Cohen, 1990), it also reveals a large amount of additional information about an individual. It informs us about a person's sex and age from its facial shape and the density and pattern of the wrinkles, a person's race by the color of the skin and the form of specific facial features. Furthermore, it allows inference about emotion and mood, degree of attention and presence to the environment, personality, and even social[7] class and professional status. It may provide information about an individual's health and degree of fatigue. It detects non-verbal communicative signs from the facial gestures, and the lip movements, inherent in the speech output and provides a visual complement to the verbal utterance. Lip-reading presents a unique glimpse of the intersection of sensory processes with modular, cognitive conjunction with heard speech (Campbell, 1992). Face perception even relieves the comprehension of the speech. Conversely, an incompatible audiovisual input can massively disturb the speech recognition, which is known as the *McGurk phenomenon* (McGurk & MacDonald, 1976). Other automatic face processes also seem to interact with each other. For instance, while same-race faces are better identified at the subordinate level, the reverse advantage is true when faces are classified by race (Levin, 1996). Additionally, the current task and the cognitive processes influence the interpretation and the processes of the lower-level features (Schyns, Goldstone, & Thibaut, 1998).

Automatic and fast processing

When you look at a familiar face, recognition seems to be automatic, thus going on without any cognitive effort. Practically speaking, you cannot look at a familiar face and decide not to recognize it (see Young & De Haan, 1995). Searle (1984) used face recognition as an example of an ability that happens 'quite effortlessly' and which he considered as not requiring complex computation, because it proceeds fast and without any significant mental energy (cf. Konorski, 1967; Fodor, 1983).

Furthermore, humans are able to identify objects very fast and accurately. Accurate identification of object pictures displayed by rapid serial visual presentation (RSVP) techniques can be revealed with presentation rates of ten pictures per second (Potter, 1976), but requires about 300 ms for further processing before the memory representation is resistant to conceptual masking from a following picture. From a pure mathematical point of view and disregarding fatigue and boredom, this rate would allow a capacity of 576,000 objects per 16-hour day (Biederman, Subramaniam, Bar, Kalocsai, & Fiser, 1999; cf. Standing, 1973). Only a third of this value would be achieved if our visual apparatus was limited to the normal scanning eye fixation rate of 3 fixations/second (Biederman et al., 1999).

Even for such highly complex objects as faces, this rate seems to be quite similar: "A familiar face is identified in about 0.5 s" (Carey, 1992, p.95). Other experiments have also demonstrated that natural objects belonging to a specified category may be classified remarkably fast (Subramaniam, Biederman, & Madigan, 2000; Potter & Levy, 1969), even when complex

[7] The social importance of a face and its expressions is particularly relevant in the Chinese culture, where 'face' does not only refer to the physical part of the head, but is comprehended as an interaction between people *per se*.

scenes had to be categorized. Such a task could be performed within 150 ms (Thorpe, Fize, & Marlot, 1996; Thorpe, Gegenfurtner, Fabre-Thorpe, & Bülthoff, 2001; Fabre-Thorpe, Delorme, Marlot, & Thorpe, 2000; Edelman & Intrator, 2002). In the subsequent studies of the present work, the recognition process is likely to be even less time-consuming (lasting no more than 100 ms). This will be demonstrated by presenting familiar faces for only 26 ms and unfamiliar faces for only 32 ms.

However, face processing is not only very fast but also highly accurate. Subjects were able to remember rather subtle aspects of the configuration of facial features to which they have earlier been exposed (Bruce, Doyle, Dench, & Burton, 1991). This high accuracy was also found when faces were degraded or presented rapidly (Biederman, 1981; Lakshminarayanan et al., 1997). Knoblich and Riesenhuber (2002) used this ability to create *effective stimuli* by simplifying stimuli step by step until there was a decrease in the firing rate for the investigated brain cell. This could be diagnostic, because a central claim of the simplification procedure is that it can be used to determine the 'dictionary of features' used to represent an object in a certain area.

In order to recognize faces so fast and accurately humans need optimised categorization processes, which break down the complexity of scene understanding into a sequence of circumscribed pattern recognition problems. One possible method to do this comes with visual attention, which allows for the possibility to focus on special points of interest (Tsotsos, 1995). Reversely, we can be 'blind' even to major aspects of natural scenes when we attend elsewhere (Li, VanRullen, Koch, & Perona, subm.).

Face encoding

For faces to be recognizable, they must be represented in the long-term memory (Tulving & Schacter, 1990). It would be naïve as well as uneconomical to assume that faces are stored as whole pictorial codes (Hochberg & Galper, 1967). They have to be encoded more structurally, because facial representation preserves information about the gray levels of the images in a size-invariant format (Bruce, Burton, Carson, Hanna, & Mason, 1994). Processing a face in terms of its physical attributes would make recognition memory more 'vulnerable' to the effects of transformation than a processing strategy emphasizing structural dimensions (Parkin & Goodwin, 1983).

Generally, experts prefer to recognize objects in their domain of expertise in a more specific way than novices do. In the field of face expertise, a *basic-level* recognition would not provide a special advantage in recognizing faces because the basic arrangement of features in a face is the same for every natural and healthy face. In every face, the eyes are paired, arranged above the nose and the mouth is the lowest part of the inner feature configuration. But the fine adjustment of the configuration and the size and form of these features, the so-called subordinate level[8], is essential for the distinction between different faces (cf. Tversky & Hemenway, 1984). Tanaka and Gauthier (1997) even define every form of object expertise as the ability to quickly and accurately recognize objects at specific or subordinate levels of abstraction.

Tanaka and Taylor (1991) and Tanaka (2001) investigated the entry level recognition in two kinds of experts: dog experts and bird experts. They found that in a speeded naming task, experts identified pictures of objects from their domain of expertise using subordinate level labels rather than basic level labels (e.g., bird experts identified a picture of a woodpecker with the label of 'woodpecker' rather than 'bird')[9]. This was validated by a *category-verification* task (Tanaka, 2001, Exp. 2) and an *identity-matching* task (Tanaka, 2001, Exp. 4). However,

[8] For a multifaceted attack on the subordinate-expertise account see Carmel and Bentin (2002).

[9] Brown (1958) pointed out that we ordinarily speak of *the* name of a thing as if there was just one, but in fact, every referent has many names on different detailed levels. It is widely believed that mental development is from the abstract to the concrete, from a lack of differentiation to increased differentiation. However, the opposite is often the case. Vocabulary often builds in the opposite direction (see also Tanaka, 2001).

when identifying objects outside their domain of expertise, the expert subjects used basic level terms (e.g., bird experts identified a picture of a beagle with the basic label of 'dog'). This pattern of results was replicated for faces by Keiji Tanaka (1996).

The role of subordinate level representations in entry point recognition was more directly shown in a category verification experiment. Tanaka and Taylor (1991) found that experts, when categorizing pictures of objects from their domain of expertise, were just as fast to categorize objects at the subordinate level as at the basic level. Based on the naming and verification measures, it can be inferred that the experts' subordinate level representations were equally accessible as their basic level representation. This experimental data suggest that entry point recognition at the basic level is not the usual one as often thought. Depending on the task demands and the amount of experience of the categorizer, the entry point can shift downward to subordinate levels and therefore demonstrate that task demands and experience have great influence on the recognition process.

Whether the level of unique identity is preferentially accessed over the basic level or whether both are accessed in parallel could be tested in a *repetition-priming* task. Ellis, Young, and Flude (1990) found that prior exposure to a face produced repetition priming of faces in a *familiarity-decision* task, but not in an *expression-* or *sex-judgment* task. They argued that a familiar face automatically activates identity level representations (in terms of subordinate-level points) and tasks that access the same identity representations (e.g., familiarity-decision tasks), which should show facilitative priming effects. According to this logic, prior exposure to a face should facilitate identity level judgments (e.g., 'Groucho Marx'), but not basic level judgments (e.g., 'human'). Jolicoeur et al. (1984) found more specific evidences, arguing against the subordinate entry level. They asked subjects to categorize pictures of common objects at super-ordinate ('furniture', 'tool', 'animal'), basic ('chair', 'saw', 'bird'), and subordinate ('kitchen chair', 'cross-cut saw', 'robin') levels of abstraction to measure the processing cost incurred by the different entry-level categorizations. Pictures were presented in either a long exposure (250 ms) or short exposure (75 ms) condition. Their main finding was that for basic level categorizations, reaction times were the same in the short and long exposure conditions. Presumably, 75 ms was a sufficient amount of time for subjects to abstract the part features, necessary for the basic-level categorizations. In contrast, subordinate-level categorizations were disrupted by the brief exposure duration, suggesting that additional visual analysis was required for subordinate-level categorizations beyond the amount required for basic-level categorizations. This is in accord with the seminal entry-level paper of Rosch, Mervis, Gray and Boyes-Braem (1976) which argues that visual classifications are initially made at a *basic-level*, and that only after such a classification, the subordinate-levels are available. As an exception, atypical members of a category were more quickly classified at their own *entry-level* rather than at the *basic-level* (Rosch, Mervis et al., 1976).

The distinction between different entry levels can also be confirmed neuropsychologically. The *middle fusiform*, the *occipital gyri* and the *parahippocampal gyri* were recruited for subordinate minus basic visual judgments, reflecting additional perceptual processing (Gauthier, Tarr et al., 2000). Furthermore, a right hemisphere advantage for face recognition based on such subordinate-level processing was observed (Rhodes, 1993).

Domain-specific memory

According to Bahrick, Bahrick and Wittlinger (1975) adults are able to recognize familiar faces with an accuracy of 90 per cent or more, even when some of those faces have not been seen for fifty years (cf. Deffenbacher, 1986). Interestingly, this remarkable ability is highly dependent on orientation (for details on this point, see section 3.1.3). Indeed, reliable processing of facial information requires a large amount of expertise (for a review see Carey, 1992) that is capable of organizing the huge data amount in a manageable way.

Twenty years ago, a positive effect of experience on memory for domain-specific information has been demonstrated in a wide range of tasks and domains (Chase & Ericsson, 1982). Expertise is not, as sometimes thought to be, a matter of general memory size, but is more correlated to a domain-specific storage. The overall memory capacity, in terms of the short-termed as well as the longer-termed, does not seem to differ between experts and novices, but with experts, the organization and the memory load for their special expertise domain is much more sophisticated (Baddeley, 1998).

Upright Expertise

Through years of practice the face recognition system becomes more specialized, but at the same time more constrained to processing the upright orientation (Schwaninger et al., in press), because of the high amount of seen upright faces. In numerous studies a disproportionate inversion effect for faces relative to various control stimuli could be shown (e.g. Scapinello & Yarmey, 1970; Yarmey, 1971; Valentine & Bruce, 1986b; Yin, 1969). The face inversion effect can be expanded to other expertise domains, like dog expertise. For instance, Diamond and Carey (1986) demonstrated that there is a similar inversion effect for dog experts when recognizing dogs. Thus, it seems that the face inversion effect is only a very strong and general phenomenon among humans, that has something to do with the high relevance of faces for humans and with the way we encounter these faces (Tanaka & Gauthier, 1997). The expertise concerning *upright faces*, but not *faces as such* could be demonstrated by a *change-detection* paradigm. Additionally, there was only an advantage for faces when faces competed with other objects for attention (Experiment 1), but this was eliminated when inverted objects were presented (Experiment 2b), and was even reversed when just a single object was presented in each display (Experiment 2a) (all experiments from Ro, Russell, & Lavie, 2001).