An investigation into the cognitive processes that mediate face perception

Isabelle Boutet

Department of Psychology
McGill University, Montréal

January, 2002

A Thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of Doctor of Philosophy

0-612-78652-8

Acknowledgements

Thanks to Avi Chaudhuri for his generous and steady support throughout my time at McGill and for his precious assistance in revising the manuscripts presented in this thesis. The Chaudhuri Vision Lab is an exceptional environment where stimulating discussions and laughs abound. Thanks to all CVL members and denizens for their friendship and for their eagerness to help with programming, experimental design, and data interpretation. Merci aussi à mes parents et à ma famille pour leur soutient moral et financier tout au long de mes études. Finally, thanks to Cary Kogan for his generosity in proof-reading the thesis and in providing constructive suggestions throughout the development of the experiments reported herein. I could not have accomplished this project without his endless ability to listen and his unconditional support.

The experiments presented in the thesis were supported by a Post-Graduate Scholarship from the National Research Council of Canada and a Doctoral Research Award from the Medical Research Council of Canada.

Abstract

A set of empirical studies is presented that examines the relationship between face perception, the modular hypothesis of cognitive function proposed by Fodor (1983), and attention. In the first study, two different manipulations were used to examine whether faces automatically trigger holistic processing operations as measured by the composite effect. The results support a modular account of face perception.

The second study introduces a novel rivalry phenomenon produced by overlapped upright tilted faces. The results indicate that this effect is dependent upon orientation with overlapped inverted faces being perceived as ambiguous in a majority of trials. The third study further examined the factors underlying this rivalry effect. It was found that contrast reversal did not influence the rivalry effect produced by overlapped upright faces and that overlapped houses did not produce rivalry. Results from both studies were taken as evidence that faces are more readily processed as Gestalts compared to other complex objects and therefore engage domain specific operations. The results also suggest that fast operations underlie perception of a face as a Gestalt. Finally, it was suggested that the rivalry effect produced by overlapped faces may illustrate informational encapsulation in face perception.

In the fourth study, faces were used to investigate the relationship between attention and modular functions. Three separate experiments showed that faces and houses compete for attention. This finding suggests that the face perception module does not have its own dedicated attentional resources but rather shares a common pool with other visual processes. Results from one experiment also suggested an advantage for faces in the allocation of attention at very short presentation times. This advantage was postulated to arise from two interacting mechanisms that is, faces capture attention over other objects and faces are more automatically encoded than other objects. Together, these studies indicate that a modular conceptualization of face processing is both appropriate and useful. They

also demonstrate the utility of faces for investigating cognitive mechanisms that mediate modular functions.

Résumé

Cette thèse décrit une série d'experiences empiriques examinant la relation existant entre la perception des visages, l'hypothèse de modularité énoncée par Fodor (1983), et les mécanismes attentionnels. Dans la première étude, deux procédures ont été utilisées pour démontrer que l'analyse holistique des visages, tel que mesuré par des visages chimériques, est automatique. En accord avec l'hypothèse modulaire, ces résultats indiquent que les visages sont obligatoirement analysés de façon holistique.

Dans la deuxième étude, un nouvel effet de rivalité produit par la superposition de deux visages inclinés est présenté. Les resultats indiquent que l'orientation des visages est cruciale pour cet effet. La superposition de visages inversés ne produit pas un effet de rivalité mais au contraire produit une perception ambigüe dans la plupart des cas. La troisième étude indique que des visages présentés en format de photographie négative produisent un effet similaire à celui obtenu pour les visages en format positive. Par contre, la superposition de maisons ne produit pas cet effet. Les résultats de ces deux dernières études suggèrent que les différentes parties d'un visage sont facilement groupées en Gestalt et donc que les visages engagent des opérations spécialisées. Les résultats suggèrent aussi que ces opérations sont rapides. Finalement, il est possible que l'effet de rivalité produit par la superposition des visages illustre une certaine impénétrabilité durant la perception des visages.

La quatrième étude examine la relation entre les ressources attentionnelles et les opérations modulaires engagées par les visages. Trois expériences démontrent que les visages et les maisons compétitionnent pour les même ressources attentionnelles. Ce résultat suggère que le module qui est responsable pour la perception des visages ne possèdent pas ses propres ressources attentionnelles mais plutôt les partagent avec d'autres processus visuels. Une des expériences démontre également que lorsque des visages sont présentés pour une courte période de temps, ceux-ci sont alors avantagés durant l'allocation des ressources attentionnelles. Cet avantage est interprété comme réflétant deux

mécanismes interdépendants —soit que premièrement les visages capturent davantage l'attention que d'autres catégories d'objets, et que deuxièment, l'analyse des visages s'avère être plus automatique que celle d'autres d'objets. L'ensemble des résultats présentés dans cette thèse supportent l'hypothèse de modularité pour la perception des visages. Ils démontrent aussi l'utilité des visages pour l'étude des facteurs cognitifs qui interviennent dans les fonctions modulaires.

Contributions of Authors

Four separate studies will be described in this thesis. At the time this dissertation was submitted, one study had been published and the results of the three other studies were being reviewed for publication. The following is a statement regarding the contributions of my co-authors to this work.

All manuscripts were co-authored with Dr. Avi Chaudhuri, who provided assistance in his capacity as research supervisor. He provided significant input in the writing and revision of the manuscripts.

Alyson Gentes-Hawn participated in the conceptualization, data collection, and data analysis of Experiment 1 of the first manuscript as an undergraduate Honors student.

Karen Borrmann contributed to the discussion and writing of the fourth manuscript. She was responsible for data collection, data analysis, and writing of Experiment 3.

Contributions to Original Knowledge

The experiments presented in this thesis provide original contributions to the study of face perception. Previous investigations in this area have led to the hypothesis that faces engage a discrete visual recognition module that is not recruited by other objects. Although, these efforts have primarily focused on the domain specificity of such a module, other aspects of a modular cognitive architecture have received little consideration. This dissertation examines this issue and provides new or improved evidence with respect to the cognitive factors that mediate face perception. In addition to providing support for a modular organization of face processing, the thesis explores the factors that govern interactions between attention and modular functions, and between modular and generalized recognition systems.

The study presented in Chapter 2 represents a first demonstration that some of the operations that underlie the processing of realistic face stimuli can be carried out with little attention.

Chapter 3 describes a novel multistable phenomenon whereby two overlapped faces cannot be perceived simultaneously. The second manuscript in the chapter indicates that overlapped non-face objects are not multistable. These studies provide further support for the claim that faces, but not other complex objects, are encoded as a Gestalt. They also suggest that the operations that underlie this type of perceptual analysis are very fast.

The study described in Chapter 4 explores the way that faces and non-face objects compete for attention. As of the date of submission, this study offered the first psychophysical demonstration that processing within the face module can be limited by allocating resources to another domain of visual recognition. Evidence suggesting that faces hold a special status in the allocation of attention is also provided.

Preface

This is a manuscript-based thesis, comprised of four self-contained research articles that are all related to a common theme – the cognitive factors that mediate face perception. Although there are separate Introduction and Discussion sections in each manuscript, a General Introduction and a General Discussion are provided at the beginning and end of this thesis, respectively. So as not to be redundant with the contents of the individual manuscripts, the General Introduction provides a general overview of our current understanding of modular cognitive architecture and of the face perception literature. The aim is to provide the reader with a broad perspective on the progress that has been accomplished in this field, the issues that remain contentious, and the questions that have yet to be investigated. Similarly, at the end of the thesis, a detailed discussion of each of the individual Manuscripts is not undertaken. Rather, the aim of the General Discussion section is to present a more general discussion of the implications of the main findings in the thesis. Current theories of modular cognitive organization are discussed in the context of face perception and possible future directions for research are presented. Finally, each chapter begins with a preface that provides a logical link between the different manuscripts.

Contents

General Introduction

List of Figures

Manuscript I

Manuscript II

Manuscript III

Manuscript IV

List of Tables

Manuscript I

General Introduction

Chapter 1

Introduction

The goal of the research described in this thesis is to further our understanding of the visual recognition system with a special emphasis on face perception. The research strategy has been to examine whether or not face processing displays the same characteristics as those that are associated with a modular organization of cognitive functions, as well as to explore some of the cognitive factors that modulate this type of architecture. As with other research endeavours, the studies in this thesis arose not only from carefully drawn out questions, but also from unexpected findings. What follows is a description of the specific issues explored in this thesis and the underlying motivation for the experiments described herein. This discussion is followed by a description of the characteristics associated with a modular cognitive architecture. A review of previous investigations into the perceptual and cognitive mechanisms that underlie face perception is then provided.

1.1 Issues Explored in the Thesis

An interpretation of the information conveyed by a face plays a vital role in social interactions among humans. Our daily experience tells us of the special status faces hold in visual perception. We are surprisingly good at recognizing and discriminating faces and expressions, despite the fine within-category comparisons required for these judgments. Because faces possess several properties not shared by other objects, it has been proposed that specialized mechanisms exist within our visual system to deal with the difficulties inherent to face perception (reviewed by Kanwisher, 2000 and Tovée, 1998). The notion that faces are processed by a special recognition system not involved in the analysis other objects is at the center of an active debate in psychology.

The bulk of the research in this area has focused on identifying differences in the way the visual system treats faces and non-face objects. A variety of empirical methodologies, including behavioural, neuropsychological, electrophysiological, and functional imaging techniques, have been employed to serve this purpose. These studies have provided ample evidence for the existence of domain specific recognition mechanisms. However, very little progress has been made in the investigation of the cognitive factors that mediate face perception. In addition to domain specificity, there are other well-defined cognitive characteristics that pertain to specialized modules and that may be associated with the module postulated to exist for faces. These include innateness, limited access to central processes, informational encapsulation, mandatory processing, fast operations, shallow outputs, fixed neuronal architecture, and specific breakdown patterns. Yet, with the exception of innateness and domain specificity, the issue of whether or not face perception displays some of the remaining characteristics has been largely unexplored. This thesis looked at this question specifically by examining the issue of whether face perception displays some of the same properties as those described by Fodor (1983) in his highly influential treatise on the modular organization of the mind.

By demonstrating that some of the modular features proposed by Fodor are indeed applicable to face perception, Chapters 2 and 3 provide additional evidence in favor of the existence of a specialized face module.

One important aspect of cognitive modularity is the *mandatory* analysis of the stimuli that the module is specialized for. The manuscript presented in Chapter 2 indicates that this type of analysis does indeed apply to face perception by showing that faces automatically trigger the operations of the face recognition system. The manuscripts presented in Chapter 3 focus on two other aspects of cognitive modularity: *informational encapsulation* and *domain specificity.* A novel rivalry phenomenon is described whereby two overlapped faces cannot be perceived simultaneously. In contrast, overlapped non-face images or contrast-reversed faces do not display rivalry. The fact that this effect is restricted to faces, occurs quickly, and is insensitive to the intentions of the observer, shows that face perception exhibits the same properties as those that have been postulated to exist for a cognitive module.

A slightly different approach is taken in Chapter 4 where faces are used to investigate aspects of cognitive modularity that are less well understood. Two questions are addressed. First, the extent to which the face module is autonomous is examined by looking at the competition between faces and non-face objects. Second, the study explores whether the deployment of attentional resources to a scene is modulated by the type of analysis triggered by competing images. The results indicate that modular processes are subject to the same attentional limitations as other visual operations. Moreover, it appears that faces hold a special status in the allocation of attentional resources. The implications of these findings for our current understanding of visual cognitive processes are discussed.

1.2 Properties of a modular cognitive architecture

In his influential treatise on the nature of cognitive processes, Fodor (1983) suggested that psychological functions are subserved by several different special purpose modules that transform incoming information into mental representations. Although the modularity hypothesis was initially applied to language, it has now been extended to other psychological processes, including face perception. The role of cognitive modules is to transmit mental representations to *central processes* that are non-modular, such as thought or problem-solving. According to Fodor (1983, 1998, 2000), cognitive modules possess four essential properties —domain specificity, encapsulation, inaccessibility, and innateness. Not all proponents of the modularity hypothesis agree that a module must display each and everyone of these properties. Chomsky (1972), for example, believes that modular processes are inaccessible, domain-specific and innate, but not necessarily encapsulated. Nevertheless, Fodor's original formulation is used in this thesis and each of the postulated properties is described below. Other properties that are derived from the four basic ones are also discussed.

1.2.1 Domain specificity

Cognitive modules are domain specific in that they operate only upon a specific type of information. A system is likely to be modular if the operations it performs are idiosyncratic. Fodor refers to eccentric stimulus domains where a highly specialized type of analysis is necessary. Therefore, the operations of a given module are closely adapted to deal with properties that are inherent to the domain in question, and that are not necessary for other domains. Different possible domains include speech perception, visual guidance of body motion, and recognition of voices and faces.

1.2.2 Informational encapsulation

The informational encapsulation property encompasses several related ideas. First, it refers to the fact that there is no cross-talk between modules. Second, modules do not have access to the operations of central processes such as expectations, intentions, desires, or beliefs. Finally, modules only consider a portion of all the information that might pertain to the domain of specialization. Therefore, by definition modules do not have access to mental facts that may be accessible to other modules. As such, transformation of a percept into a mental representation is accomplished in isolation of other modular and central processes. Perceptual illusions are a compelling illustration of informational encapsulation. For example, despite having knowledge about the structure of the Ames room, one still does not view the person on the taller side of the room as the same size as the person on the other side.

Informational encapsulation provides two important advantages: speed and a reduced memory load. Because a given module ignores all but immediately relevant information, the operations it performs are fast. Moreover, because only a small portion of all the available input needs to be searched, the memory space required to match relevant information with stored representations is minimized. Although Fodor (1983) clearly states that access to information is constrained in a cognitive modular architecture, to what extent modules are autonomous is less well defined. The author mentions that the 'widely accepted picture' is that modules compete for *central resources* such as short-term memory and attention. Allocating such resources to the operations of a given module results in a decrement in the performance of others. Chapter 4 addresses this less well known aspect of modularity in further details.

1.2.3 Inaccessibility

Just as modules have no access to feedback from central processes, central processes have no access to the ascending levels of representations that a module generates. Only the final output of a module is available to those cognitive processes that underlie voluntary overt behaviors. Therefore, operations and intermediate representations that a module generates are inaccessible to central processes and unavailable for explicit report. Inaccessibility may be the result of lower levels of analysis being discarded 'on the fly' with only high-level representations being transferred into memory. This allows for a fast and efficient analysis of the information before more elaborate representations are transmitted to central processes. Fodor (1983) cites anecdotal evidence for this property. For example, one can tell the time from looking at a watch despite being unable to recall the shape of the numerals used to represent the time.

1.2.4 Innateness

The information and operations proprietary to a module are innately specified, or "genetically preprogrammed". Innateness refers to the idea that the neural mechanisms responsible for the operations of a module are already present in the newborn infant. Innateness also means that these mechanisms develop according to specific, endogenously determined patterns shaped by the impact of environmental factors.

1.2.5 Other properties

Several additional characteristics are associated with, or arise from, the four basic aforementioned properties. First, analysis of relevant information by a module is mandatory because the module provides the only route whereby central processes can gain access to specific mental representations. Mandatory processing implies that the operations of the module are automatically triggered by the stimulus that they apply to. It also means that the module performs its computation irrespective of the desires of the subject. For example, you cannot hear speech as noise even if you so desired.

Second, encapsulation and mandatory processing bestow speed to modular processes. Modules are fast not only because feedback from central processes and other modules is minimal, but also because automatic operations require little computation. Given that modules are designed to perform highly specialized operations, the complexity of these operations does not hinder the speed of execution.

Third, modules are associated with a fixed neural architecture. For example, there are specific neuronal substrates in the occipital and temporal lobes that underlie perception and language. This implies that particular patterns of cognitive deficits should arise from damage to different modules. The innateness property stipulates that these substrates are present at birth and develop in a specific pattern under appropriate environmental influences.

Finally, the encapsulation property implies that modules deal with limited information and as such, that the outputs of modules must be shallow. Because there is no access to prior knowledge and to information available to other modules, the end result of modular analysis is restricted. Visual modules, for example, deliver 'basic' categories that are ready for use by the central cognitive systems. A stimulus may be classified as a *bird* at the modular level, but further interactions with central processes and other modules would be required to identify it as a *sparrow*. It is important to note that the terms shallow and basic are only relevant within the domain of specialization of the module. For example, a face recognition system may derive mental representations that are shallow with respect to that domain but yet correspond to a higher level of within-category classification than that derived by a general recognition system.

1.2.6 The modularity hypothesis

The modularity hypothesis proposed by Fodor (1983) involves a set of pre-defined properties and derivative characteristics. As Fodor stresses, these characteristics serve as 'landmarks' that can be used to investigate whether or not particular psychological functions are modular. A similar approach is adopted in this thesis. The results from the three manuscripts are considered in light of whether face processing displays the same characteristics as those that are associated with a modular organization of cognitive functions. In the last manuscript presented, faces are used to investigate aspects of cognitive modularity that are less well understood.

It is important to note that modularity is one hypothesis that has been proposed to account for the underpinnings of the mind. Other modular accounts of cognition as well as entirely different models based on functional specificity rather than domain specificity also exist. Because this thesis draws a parallel between Fodor's (1983) modularity hypothesis and face perception, these alternative views are omitted from the discussion. This does not mean that one should neglect to critically examine Fodor's model, but rather that the model provides a theoretical foundation for the issues that are explored in this thesis. Whereas there are some features of the modularity hypothesis that have survived the test of experimentation, other aspects remain undetermined. For example, the fact that neurons in the early stages of visual processing, such as those found in area V1, are susceptible to attentional modulations (Motter, 1993) suggests that the extent to which modular operations are encapsulated may need to be reconsidered. Another related example stems from the finding that low-level perceptual phenomena, such as the motion aftereffect, can be influenced by task demands (Chaudhuri, 1990; Mukai & Watanabe, 2001). This thesis is not meant to address the debate on 'how the mind works'. Rather, it provides evidence that face perception can be classified as a modular process and consequently uses faces to further examine issues that are less well defined in the modularity hypothesis.

1.3 Review of Literature

The review of the literature is organized in a fashion that permits an examination of the past research on face perception in the context of the different characteristics that are associated with modular cognitive systems. Most studies in this area have focused on the innateness and domain specificity aspects of modularity and as such both topics will be reviewed extensively. Findings that are relevant to the 'fixed neuronal architecture' property are also examined. Finally, investigations that have a bearing on the issue of mandatory processing are also reviewed. Whether face perception displays additional modular properties will be further discussed in the general conclusion.

1.3.1 Face perception and domain specificity

Domain specificity entails the existence of unique operations that are specialized for dealing with the unique characteristics of a particular type of stimulus. As was first suggested by Teuber (1968) and later by Fodor (1983), face recognition may be a perfect candidate for this type of analysis. Indeed, it would not be surprising if the important role that identity plays in social interactions, together with the difficulties inherent to discriminating between highly homogenous faces, were sufficient conditions to warrant modularization within a general visual recognition system. Therefore, meeting the domain specificity criterion requires that the mechanisms that underlie face recognition operate differently from either a generalized recognition system, or from other domain specific recognition systems. Support for this notion can be found in behavioral, neuropsychological, electrophysiological and functional imaging studies.

1.3.1.1 Behavioral studies

Yin (1969) reported the first behavioural evidence that faces and objects are processed differently. He found that upside-down faces are disproportionately more difficult to recognize than upside-down objects such as airplanes, buildings, and costumes (reviewed by Farah, Tanaka & Drain, 1995a; Valentine, 1988). This so-called *inversion effect* is robust across different experimental paradigms and has been replicated for a variety of stimulus types. Yin took the inversion effect as evidence that faces and objects are recognized using different encoding strategies. Specifically, he argued that face recognition relies more heavily on holistic encoding than on featural encoding, while the opposite is true for object recognition. The greater disadvantage for recognition of inverted faces would therefore arise because holistic information is more difficult to extract than featural information in inverted stimuli. Whereas this distinction is now widely accepted among face researchers, it is conceptualized in several different ways (reviewed by Farah et al, 1998).

Some researchers have proposed that faces are perceived as Gestalts (Bradshaw & Wallace, 1971; Ellis, 1975). Sergent (1984) provided support for Gestalt face encoding by showing that the features of upright, but not inverted, faces are perceived in an interactive manner.

Rhodes and colleagues (Rhodes, 1988; Rhodes, Brake & Atkinson, 1993) use the terms first- and second-order features to refer to the different types of information used to recognize upright and inverted faces. First order features are discrete features such as the eye, nose, chin, etc. Second order features are configural properties such as the spatial relations between the first order features and their positions. By examining how changes in the position and orientation of facial features affect perception of upright and inverted faces, Rhodes and colleagues have found evidence that coding of second-order relations, or *configurations*, is more susceptible to inversion than is coding of first-order features.

Similarly, Diamond and Carey (1986) have proposed that faces contain featural as well as first- and second-order relational information. First-order information consists of spatial relations between the parts such as the nose being below the eyes and above the mouth. Second-order relations describe the relative size of these spatial relations with respect to a face prototype. Whereas first-order relations are used to differentiate a face from other objects, second-order relations, or *relational* information, are used to differentiate one face from another. A number of studies have provided evidence that the face inversion effect arises from a disruption in relational information (Leder & Bruce, 2000; Carey & Diamond, 1994; Hole, 1994; Searcy & Bartlett, 1996; Young, Hallewell & Hay, 1987). Leder and Bruce, for example, have shown that the face inversion effect is more pronounced for faces that are unique with respect to relational information than with respect to other types of information.

Farah and colleagues (Farah et al., 1995a, 1998; Tanaka & Farah, 1993) argue that face recognition differs from other types of recognition in that it involves relatively little part decomposition, meaning that facial features are represented in a *holistic* fashion. Support for holistic face processing comes from the finding that recognition of a face part is superior when the part is presented in the context of the face than in isolation (Tanaka & Farah, 1993), and from the finding that changes in a given facial feature affects recognition of other features (Tanaka & Sengco, 1997). Houses and inverted faces and do not produce the same results, suggesting that holistic representations are more important for recognition of upright faces than for recognition of inverted faces and non-face objects.

Despite the use of different terminologies, all of these models emphasize the importance of the overall structure of the face over its discrete features. It is important to note that this does not imply that faces and non-face objects are processed in a qualitatively different manner. Rather, it is generally accepted that these two stimulus categories are recognized using information at opposite ends of a continuum between part-based and holistic representations. Chapter 3 provides further evidence that inversion disrupts encoding of faces as Gestalts but has little influence on the perception of non-face objects.

Contrast reversal is another manipulation that disproportionately affects face recognition as compared to recognition of other objects (Bruce & Langton, 1994; Galper, 1970; Gauthier, Williams, Tarr & Tanaka, 1998; Hayes, 1988; Johnston, Hill & Carman, 1992; Kemp, McManus & Pigott, 1990; Liu & Chaudhuri, 1997). This so-called *photographic negative effect* may reflect the important role surface properties, such as pigmentation and shape-from-shading, play in face recognition (Bruce & Langton, 1994; Hayes, 1988; Kemp et al., 1996).

Other manipulations have been shown to have greater effects on face recognition than on the recognition of non-face objects. However, the factors underlying these differences are less well understood. These include variations in lighting direction (Johnston et al., 1992), and detection superiority effects (Homa, Haver & Schwartz, 1976; Purcell & Stewart, 1988). These behavioural studies indicate that faces exhibit functional characteristics that are not found in object recognition and as such provide further support for domain specificity in face recognition.

1.3.1.2 Neuropsychological studies

Neuropsychological studies provide the most compelling evidence that distinct mechanisms are involved in the recognition of faces. Numerous cases of prosopagnosic patients who have lost their ability to recognize familiar faces but can still identify people on the basis of voices or gait have been reported (reviewed by Young, 1992). Consistent with the notion that faces engage holistic encoding processes, prosopagnosic patients appear to be unable to perceive faces in a unified fashion despite being able to describe and identify individual features in a face (Saumier, Arguin & Lassonde, 2001).

Prosopagnosia is often associated with recognition deficits for other object categories. In view of this co-morbidity, it has been suggested that the deficit rests in an inability to distinguish between highly homogeneous complex stimuli rather than faces *per se* (Damasio, Damasio and Van Hoesen, 1982). However, reports of 'pure' cases of prosopagnosia where patients cannot recognize faces but are still able to distinguish between different exemplars of other highly homogeneous object categories, such as cars (Sergent & Signoret, 1992), sheep (McNeil & Warrington, 1993), or dogs (Bruyer et al., 1983), shed doubt on this proposal.

Perhaps the most striking evidence in favor of a segregated face module comes from patient CK who is severely impaired at reading and object recognition and yet displays normal face recognition ability (Moscovitch, Winocur & Behrmann, 1997; Moscovitch & Moscovitch, 2000). Interestingly, CK has difficulties recognizing inverted and 'fractured' faces where facial parts are separated by gaps such that their spatial configuration is altered. Conversely, prosopagnosic patients are better than normals at matching inverted faces (Farah, Wilson, Drain & Tanaka, 1995b). Double dissociations between face and object recognition strongly support the existence of a domain specific face module. Indeed, neuropsychological studies have an advantage over other investigative approaches in that cortical mechanisms that are *necessary* for a given function can be identified. Although they provide corroborating evidence for the existence of specialized face mechanisms, the electrophysiological and brain imaging studies reviewed below do not afford this advantage. They nonetheless provide corroborating evidence for localized neuronal substrates involved specifically in face perception.

1.3.1.3 Electrophysiological studies

The presence of cells in the inferior temporal cortex (IT) of monkeys that respond selectively to faces was first reported in 1972 (Gross, Rocha-Miranda & Bender). Since then, numerous investigations have replicated this finding in various primate species and even sheep (reviewed by Gross, 1992 and by Perrett, Mistlin & Chitty, 1987). Cells that respond significantly more to faces than to other equally complex stimuli can be found in the superior temporal sulcus (STS) of area IT and in several other areas that have synaptic connections with IT, including the amygdala and the frontal lobe. Cells in IT and STS may respond best to a particular view of a face or they may be viewpoint invariant. Whereas some of these cells only respond to complete faces, others show a decline in response when facial features are removed, or respond to features in isolation. Moreover, the responses of IT and STS cells are also sensitive to changes in orientation, direction of illumination, expression, and gaze direction. Finally, the response profile of face cells supports the importance of configural information in face processing —these cells will not respond to a jumbled arrangement of facial features.

Whereas the selectivity of cells in IT is suggestive of specialized mechanisms that distinguish faces from other objects, whether these cells are actually involved in the recognition of faces remains open to debate. The fact that lesions in area IT do not produce the human equivalent of prosopagnosia in monkeys suggests that this is not the case. However, this evidence is complicated by difficulties inherent to designing comparable tests for this deficit in monkeys and humans (Heywood & Cowey, 1992). Regardless of whether IT cells are involved in recognition or perception, their selectivity and organization support the existence of specialized face selective mechanisms. Indeed, faces appear to be the only class of visual object to have cells that respond to them specifically. In contrast, complex objects are represented as a pattern of responses across various columns of feature selective cells (Tsunoda, Yamane, Nishizaki & Tanifuji, 2001).

1.3.1.4 Functional imaging studies

The use of fMRI (functional magnetic resonance imaging) technology has shown that activity in the right fusiform area, which is located in human area IT, is stronger during passive viewing or discrimination of different individual faces as compared to letter strings (Puce, Allison, Asgari, Gore & McCarthy, 1996), common objects (Kanwisher, McDermott & Chun, 1997; McCarthy, Puce, Gore & Allison, 1997), hands (Kanwisher et al., 1997), animals (Kanwisher, Stanley & Harris, 1999), buildings (Epstein & Kanwisher, 1998), and scrambled faces (Clark, Maisog & Haxby, 1998; Puce, Allison, Gore & McCarthy, 1995). This region, referred to as the fusiform face area (FFA), generally corresponds to the locus of damage in prosopagnosia and may thus constitute the neural correlate for a face module. Whether the FFA corresponds to areas IT and STS of the monkey remains speculative.

Further support for a modular organization of the visual recognition system comes from the finding that discrete brain regions located in the vicinity of the FFA are activated by presentation of other stimulus categories, such as chairs (Ishai, Ungerleider, Martin, Schouten & Haxby, 1999), houses (Aguirre, Zarahn & D'Esposito, 1998; Ishai et al., 1999), and places (Epstein & Kanwisher, 1998). These findings complement reports of patients with visual recognition deficits restricted to non-face object categories (e.g., Moscovitch et al., 1997).

Despite considerable evidence that activity in the FFA is domain specific, some studies suggest that this region also responds significantly, or with equal strength, to other object categories such as cats (Tong, Nakayama, Moscovitch, Weinrib & Kanwisher, 2000) and animals without faces (Chao, Martin & Haxby, 1999; but see Kanwisher et al., 1999). Also surprising is the finding that activation in the face area is not affected by inversion (Haxby, Ungerleider, Clark, Schouten, Hoffman & Martin, 1999; Kanwisher, Tong & Nakayama, 1998). Rather, turning faces upside-down appears to increase responses in regions selective for non-face objects. This finding suggests that the face inversion effect does not arise from a failure by inverted faces to recruit face-specific mechanisms, but rather from inverted faces recruiting general object recognition processes. Finally, it has been argued that objects produce widely distributed and overlapping activation in human area IT such that different object categories may be represented on the basis of specific response patterns rather than activation of a specific location (Haxby, Gobbini, Furey, Ishai, Schouten & Pietrini, 2001).

Perhaps the strongest argument against the notion that the FFA is a domain specific face module comes from the finding that activation in this region is modulated by expertise. Gauthier and colleagues (Gauthier, Skudlarski, Gore & Anderson, 2000) have found that FFA responses to birds and cars are greater in expert subjects than in non-experts. Responses in the FFA have also been shown to increase with increased expertise in discriminating novel stimuli called 'greebles' (Gauthier, Tarr, Anderson, Skudlarski & Gore, 1999a). These results have lead Gauthier and colleagues to suggest that the FFA is specialized for visual expertise rather than face processing. However, the fact that activation in the FFA was strongest for faces irrespective of expertise, and that greebles are facelike in several different aspects calls into question this conclusion. The expertise hypothesis is further discussed below.

1.3.1.5 The role of expertise

Behavioural, neuropsychological, electrophysiological, and brain imaging studies have provided corroborating evidence that faces engage different processing operations and cortical mechanisms than those engaged by other complex objects. However, the extent to which face processing is domain specific remains disputed because these operations and mechanisms may not be unique to faces. Gauthier and colleagues (Gauthier & Tarr, 1997; Gauthier et al., 1998; Gauthier, Behrmann & Tarr, 1999b; Gauthier, et al., 1999a; Gauthier, Skudlarski, Gore & Anderson, 2000) have suggested that most of the phenomena that distinguish faces from other objects can instead be accounted for by expertise.

Evidence in favor of the expertise hypothesis first came from the finding that expert dog judges are subject to the same inversion effect as that observed with faces (Diamond & Carey, 1986). Gauthier and colleagues have provided further support for the expertise hypothesis by showing that observers who are trained at discriminating novel homogeneous stimuli called 'greebles' can display inversion, relational, and photographic negative effects (Gauthier & Tarr, 1997; Gauthier et al., 1998). Gauthier and colleagues have also shown that activation in the FFA can be modulated by expertise using both greebles (Gauthier et al., 1999a) and real-life stimuli (Gauthier et al., 2000). Finally, there is evidence that activation in the FFA is superior for same-race faces because we presumably have more experience with them as compared to different-race faces (Golby, Gabrieli, Chiao & Eberhardt, 2001).

The evidence provided by Gauthier and colleagues has been criticized on the basis of two grounds. First, greebles are face-like or human-like in many aspects and as such may recruit face-specific mechanisms. They are composed of three rounded parts —a base, body and head— one on top of each other with protrusions that may be labeled penis, nose and ears (Biederman & Kolacsai, 1998). Moreover, their protrusions are arranged in a face-like manner with two horizontally displaced parts being arranged symmetrically above two vertically displaced parts. Finally, observers are trained to identify greebles with names, a condition that may encourage a human interpretation (Kanwisher, 2000). Second, the face-specific behavioral effects tested with greeble experts were only partially replicated, or failed to be replicated (Gauthier & Tarr, 1997; Gauthier et al., 1998). For instance, the inversion effect, which is generally measured on the basis of accuracy (Valentine, 1988), was only replicated with reaction times

(Gauthier & Tarr, 1997). In view of these shortcomings, the expertise hypothesis put forward by Gauthier hypothesis may need to be reconsidered.

Also relevant to the expertise hypothesis are studies on face recognition training. Indeed, if the face module is responsible for expert within-category discriminations, then one would expect that recognition of faces could improve with appropriate training, assuming that there is room for such an improvement. To address this question, Malpass (1981) examined whether disadvantages in different-race recognition can be eliminated with training. His findings indicate that recognition training can only produce short-term improvements in the recognition of other-race faces, and no improvement in the recognition of own-race faces. However, it is possible that recognition of other-race faces is already at ceiling, even if it is lower than that of own-race faces. The training method used may also not have been appropriate for producing long-term improvements.

There is some indication from electrophysiological recordings that training on a set of highly homogenous stimuli increases the proportion of IT neurons that respond to such stimuli (Kobatake, Wang & Tanaka, 1998). Moreover, it appears that such training may lead to cellular responses that are similar to those produced by faces in that preserved configuration is essential if the cell is to respond to the novel object (Logothetis, 2000). Even though this finding is consistent with the notion that area IT is a visual expertise area, there is no evidence that the neurons recorded belonged to the same population as those involved in face processing and in fact, the trained cells were not consistently responsive to faces in the study by Kobatake and colleagues.

The evidence in favor of the expertise hypothesis does not preclude the existence of a segregated cortical locus that carries out operations specialized for face perception. Indeed, even if the FFA is devoted to discriminating any visually homogeneous stimulus for which expertise has been developed, it may nonetheless be the case that its operations are specialized for face recognition because faces are the only object category that meets the sufficient, but not necessary, criteria to recruit this system. Therefore, the existence of a module that specializes in face processing seems highly plausible, irrespective of whether the domain of specialization of this module is face perception *per se* or expertise for within-category discriminations that necessitate the use of configural information.

1.3.1.6 Conclusion

Past research has provided converging evidence for domain specificity in face processing using a variety of empirical techniques. These studies indicate that faces trigger processing mechanisms that are not implicated in the analysis of other visual objects. Furthermore, it appears that the operations applied to faces are closely adapted to deal with difficulties inherent to their recognition. Specifically, faces are thought to be differentiated on the basis of holistic and configural information, probably because featural information is generally of little use for discriminating between exemplars of a highly homogeneous stimulus class.

It is important to note that Fodor's (1983) description of domain specificity does not state uniqueness as a mandatory condition. Rather, it stipulates that the processes that are available for the module's proprietary domain are not available to other cognitive abilities. With the exception of a few studies, the evidence reviewed above clearly indicates that faces trigger specialized functional mechanisms that do not apply to other equally complex stimuli. Evidence from neuropsychological studies also indicates that distinct cortical mechanisms perform the operations necessary for face recognition. Therefore, faces appear to meet modular property of domain-specificity in that they engage special operations that do not apply to other object categories. The question remains open as to whether or not these operations are unique to faces or can extend to other expert visual discriminations.

1.3.2 Face perception and innateness

The innateness property of modularity implies two conditions: first, that some of the functions that the module carries out are functional at birth, and second, that the development of the module follows a 'genetically preprogrammed' course given the presence of appropriate environmental influences.

Results from a number of studies suggest that shortly after birth, faces are preferred over other types of images (Goren, Sarty & Wu, 1975; Johnson, Dziurawiec, Ellis & Morton, 1991), and familiar faces are preferred over non-familiar ones (Bushnell, 1982; Walton, Bower & Bower, 1992). For example, infants as young as one-hour-old are more accurate at following a schematic face with their heads than a scrambled face (Johnson et al., 1991). Although these findings are consistent with an innate face recognition system, whether neonates respond to faces *per se* or to some more fundamental characteristic of the stimulus, such as phase or spatial frequency, is still unclear. Compelling evidence for innate face recognition mechanisms comes from the report of a patient who sustained brain damage at postnatal day one (Farah, Rabinowitz, Quinn & Liu, 2000). This patient displays the same lesions and behavioural deficits as those associated with adult-acquired prospopagnosia, suggesting that the neuronal networks that are necessary for face recognition are present at birth.

Face processing abilities undergo development until they reach their full capacity at approximately 12 years of age, as indicated by the finding that 12 year olds do better than 6-year-olds, but no worse than adults, on face perception and recognition tasks (Ellis, 1992). The processes underlying this improvement in face processing have yet to be determined. Carey and Diamond (1977) have argued that between the ages of six and ten, face perception is subject to an 'encoding switch' from featural to configural encoding strategies. Accordingly, they have found that 6-year olds' recognition performance is more affected by changes in accessories (e.g., eye glasses and hairstyle) and facial expression than that of older children. However, there is evidence that six year old children display the same advantage for recognition of a face part presented in the context of a whole face as that recorded in adults (Tanaka, Kay, Grinnell, Stansfield & Szechter, 1998). It has also been shown that six year olds are subject to the composite effect (Carey & Diamond, 1994). This effect is defined as better recognition of two half faces when presented in a non-composite misaligned condition than when presented in a composite aligned condition. Poor recognition of composites is believed to reflect holistic face encoding whereby the new holistic representation formed by the aligned halves interferes with recognition of its constituent parts. These findings are in disagreement with the encoding switch hypothesis and instead suggest that some of the processes that distinguish faces from other objects are operative by age six. The results of a recent study by LeGrand, Mondloch, Maurer and Brent (2001) are consistent with this conclusion. They examined whether configural face encoding is present in patients whose congenital cataracts had been removed at least nine years before testing. It was found that deprivation of patterned visual input from birth until 2-6 months of age results in permanent deficits in the development of configural face processing abilities.

Thus, face processing abilities may be genetically determined with faces being preferred over other types of stimuli early after birth, and face-specific operations being present at a young age. This suggests that faces hold a special status in visual processing even when little or no experience has been acquired. It is important to note that the question of whether or not face-specific operations are innate is independent of the question of domain specificity. As such, those who support the existence of specialized face recognition mechanisms may not agree that these mechanisms are present at birth. The literature reviewed herein suggests that face processing displays both of these characteristics.

1.3.3 Face perception and fixed neuronal architecture

The double dissociation between prosopagnosia and agnosia for other complex objects, the existence of face selective cells in a circumscribed region of the inferior temporal cortex, and the selective activation of the FFA to faces, all point to the existence of a fixed and segregated neuronal architecture for face processing. In a similar vein, there may be a discrete cortical locus for face learning. Tippett, Miller and Farah (2000) describe a novel case of impairment in learning to recognize new faces, which they termed 'prosopamnesia'. Patient CT has spared ability to recognize faces learned before his brain injury, preserved face perception, and preserved memory for other visually complex material. However, CT cannot learn new faces encountered after his operation. The existence of prosopagnosia also demonstrates that the face perception system exhibits specific breakdown patterns, a characteristic that is derived from the idea that hardwired neurological structures underlie modular operations.

1.3.4 Face perception and mandatory processing

Mandatory processing stipulates that the operations of a module are automatically triggered by relevant stimuli. However, this does not mean that modular processes should be unaffected by attentional manipulations. Rather, the module will still operate in its idiosyncratic way without attention, but access to the processes that mediate storage and report will be disrupted. Brain imaging, behavioural, and neuropsychological studies that have examined this issue are reviewed below.

1.3.4.1 Functional imaging studies

Brain imaging studies indicate that a variety of attentional manipulations can enhance or suppress activation in face-selective areas (Clark at al., 1997; Haxby et al., 1994; O'Craven, Downing & Kanwisher, 1999; Vuilleumier, Armony, Driver & Dolan, 2001; Wojcuilik, Kanwisher & Driver, 1998). However, whether diverting attention away from a face disrupts the *way* that it is encoded has received little consideration. Eimer (2000) recorded N170 ERPs (event-related brain potentials) to examine whether face-specific operations are modulated by attention. N170 ERPs are interesting because they are believed to reflect an early structural face encoding process. Indeed, N170 ERPs are elicited by faces but not cars, hands, furniture, or scrambled faces. Moreover, N170 components to faces are delayed by about eight milliseconds when a face is inverted. Eimer found that diverting attention away from a face further delays the onset of N170 ERPs. The author took this finding to indicate that structural face encoding is affected by attentional factors. However, the fact that N170 components were also delayed when attention was diverted away from inverted faces suggests that the effect may not be face-specific.

1.3.4.2 Behavioural studies

The visual search paradigm provides an interesting method for investigating whether holistic face encoding is automatic. Visual search studies with simple stimuli indicate that targets that differ from distractors by only one feature pop out. However, targets that differ from distractors by a conjunction of two or more features do not. These results suggest that simple features are detected pre-attentively, but that the binding of stimulus features into a holistic representation requires attention (e.g., Treisman, 1988; Treisman & Gelade, 1980). Results from visual search studies indicate that line-drawn and digitized faces do not pop out from inverted and scrambled faces (Kuehn & Jolicoeur, 1994; Nothdurft, 1993), and that a happy face icon does not pop out from sad face icons (Suzuki & Cavanagh, 1995), suggesting that properties that are inherent to faces are not automatically encoded. Nonetheless, speed of detection has been shown to be superior for faces as compared to other objects (Purcell & Stewart, 1988), suggesting that faces hold an advantage in either speed of processing or amount of attention required for processing. This advantage is particularly salient when searching for one's face amidst the faces of strangers (Tong & Nakayama, 1999).

Although facial features do not pop out, the emotional content of faces is detected pre-attentively with sad, happy, or threatening faces popping out amongst emotionally neutral faces (Öhman, Lundqvist & Esteves, 2001; White, 1995, but see Fox et al., 2000; Purcell, Stewart & Skov, 1996 for a lack of pop-out effect). However, this pop-out effect does not disappear after inversion, suggesting that face-specific operations may not be responsible for the fast and efficient search recorded.

Three psychophysical experiments have directly examined whether attention influences encoding of the global properties of a face. Reinitz, Morrissey and Demb (1994) compared recognition of intact faces identical to those shown at learning with that of conjunction faces created by combining features from two different faces. Devoting attention to faces at learning produced a greater performance in recognition for intact faces than for conjunction faces. In contrast, distracting attention away from faces at learning significantly impaired recognition of intact faces but did not affect recognition of conjunction faces. These results suggest that attention is necessary for encoding the spatial relations between facial features. However, the use of line-drawn stimuli by Reinitz et al. sheds doubt on the generality of this conclusion because their processing depends less heavily on holistic information than that of photographs (Leder, 1996). Furthermore, Reinitz, Bartlett and Searcy (1997) later found that dividing attention produced a larger performance decrement for faces whose features had been altered than for faces whose configuration had been altered. This finding is in direct contradiction to the results reported by Reinitz et al. (1994).

Finally, Palermo and Rhodes (2001) asked participants to match two faces presented on each side of a target face during learning. Recognition of parts of the target face was measured either in isolation or in a face context. The advantage for recognition of a face part when presented in a face context was found for attended faces only, suggesting that holistic face encoding is influenced by attention. However, these results may reflect a limitation in the number of faces that can be holistically encoded simultaneously, rather than a failure for faces to trigger holistic processes in the absence of focal attention. Indeed, finding a whole/face advantage in the divided attention condition would have required that all three faces presented during encoding be holistically processed.

1.3.4.3 Neuropsychological studies

Vuilleumier and colleagues (Vuilleumier, 2000; Vuilleumier & Schwartz, 2001) have provided evidence that faces may be encoded more automatically than non-face objects. Patients suffering from unilateral neglect can detect an image presented in their contralesional hemifield, but fail to do so when a competing stimulus is presented on the ipsilesional side. This deficit is believed to reflect an inability to direct attention toward contralesional space in the presence of competing information. Vuilleumier and colleagues have shown that these patients are less likely to miss a face than a scrambled face or object, and a happy or angry face than a neutral face. This finding suggests either an advantage for faces in capturing attention over other types of stimuli, or that the operations that underlie face processing are automatic.

1.3.4.4 Conclusion

Studies that have examined the influence of attention on face processing have not adequately examined whether the operations of the face module are mandatory in several important ways. First, brain imaging studies provide no indication that face-specific operations can be disrupted by withdrawing attention, they only show that activation in face selective areas can be modulated by attention. Second, behavioural studies that directly examined whether attention is necessary for the deployment of face-specific operations are limited because they produced inconsistent results (Reinitz et al., 1994; 1997), and because the procedure employed may not have tapped into focal attention mechanisms (Palermo & Rhodes, 2001). Finally, the studies by Vuilleumier et al. (2000, 2001) offer two equally plausible possibilities —that faces are more automatically encoded than non-face objects, or that faces attract more attention than non-face objects. The experiments presented in Chapter 2 address some of these limitations and provide evidence that holistic face encoding processes are mandatory.

1.4 Summary

This chapter has reviewed the major tenets of Fodor's (1983) modularity hypothesis and studies that have addressed the question of whether face perception displays modular characteristics. Modules exhibit four essential properties —domain specificity, innateness informational encapsulation, and inaccessibility. Additional modular properties that are derived from these four basic characteristics include mandatory operations, speed, a fixed neural architecture with specific breakdown patterns, and shallow outputs.

Past research on face perception suggests that face processing displays three important modular characteristics —domain specificity, innateness, and a fixed neuronal architecture with specific breakdown patterns. First, there is evidence that specialized operations are applied to faces but not to other complex objects. Second, the finding that some face perception abilities are operative at birth or in early childhood suggests that face-specific processes may be innately specified. Finally, imaging, electrophysiological, and neuropsychological studies have provided support for the existence of an anatomically segregated neuronal substrate in the right fusiform gyrus of the inferior temporal cortex that preferentially responds to faces. The existence of a double dissociation between prosopagnosia and agnosia for other complex objects indicates that this region is essential for face recognition and that impaired processing within the face module is associated with specific breakdown patterns. Whether the domain of specialization of this region is face perception, or expertise for any class of homogeneous stimuli whose recognition requires within-category discriminations, remains disputed.

Other studies have provided evidence that the deployment of face-specific operations may require attention, contradicting the notion that modular processes are mandatory. However, the experimental procedures that have given rise to these results are faced with criticisms that warrant further experimentation. Previous research on face processing is also limited in that other modular properties have remained largely unnoticed, including informational encapsulation, inaccessibility, speed of processing, and shallow outputs. The thesis sought to address some of these limitations by providing improved or novel evidence that further links modular properties with face processing.

The results presented in Chapter 2 suggest that face-specific operations are mandatory in that holistic encoding operations are automatically applied to faces. The results presented in Chapter 3 indicate that overlapped upright faces, but not overlapped inverted faces and houses, produce perceptual rivalry. This finding provides additional support for domain specificity in face perception by showing that Gestalt grouping principles are more readily applied to faces than to other complex objects. The results also suggest that processing of faces as Gestalts is fast. The possible relationship between this rivalry effect and informational encapsulation is also considered.

The findings presented in Chapters 2 and 3 provide further evidence that face processing exhibits modular properties. In Chapter 4, faces are used to investigate aspects of a cognitive modular architecture that are less well understood. The results suggest that the face module does not have its own dedicated attentional resources but rather shares a common pool with other visual recognition mechanisms. Evidence for an attentional face advantage is also reported. The thesis demonstrates the value of using the modularity hypothesis as a foundation for investigating visual perception mechanisms and the cognitive processes that mediate this function.

Experimental Findings

Preface

The manuscript provided in Chapter 2 examines whether the operations of the face module are automatically applied to the module's proprietary stimuli. There is evidence that faces can be distinguished from other complex objects in that their perception and recognition relies more heavily on the global characteristics of a face (also termed holistic, relational, or configural information) than on featural information. The work is motivated by the assumption that this type of analysis should be applicable to faces irrespective of the allocation of attention if processing within the face module is mandatory. This issue was examined using a previously reported test of holistic face encoding, the composite effect. Two separate experiments are presented, both of which support the idea that the global structure of a face can be encoded even if the face is not attended to. These results lend support to the notion that face-specific processes are automatic, a feature that is consistent with the modular architecture postulated by Fodor (1983).

Chapter 2

Manuscript I

The Influence of Attention on Holistic Face Encoding

Boutet, Gentes-Hawn & Chaudhuri. Submitted to *Cognition*.

I.1 Abstract

We examined the influence of attention on the holistic nature of face representations using the composite effect (Young et al., 1987). In Exp. 1, stimuli composed of a face superimposed on a house were shown during encoding, Ss delineated either the face or the house, thus manipulating attention away or toward the face. In Exp. 2, an intact face image was presented with letters scrolling from top to bottom. Ss were asked to either ignore the letters or read

them and decipher the words that they formed. Aligned and misaligned composite stimuli were shown at testing. We found recognition performance to be consistently better for misaligned than aligned stimuli, regardless of the allocation of attention during encoding. The composite effect was therefore present in both attended and unattended faces. We take this as evidence that holistic encoding is one aspect of face analysis that can be performed without attention.

I.2 Introduction

It is widely accepted that faces and objects engage different processes that may be performed by distinct brain areas (see reviews by Biederman & Kolacsai, 1998 and Tovée, 1998). One well established difference between faces and objects concerns the type of information used for their recognition. Different experimental paradigms have provided converging evidence that holistic information is more important for face than for object recognition[1]. The basis for this dissociation is, however, disputed. On one hand, mechanisms specialized for dealing with the difficulties inherent to face recognition may exist because of the important role faces play in human interactions (e.g. Farah, 1992; Farah, Wilson, Drain & Tanaka, 1995). Alternatively, faces may engage an expert subordinate-level recognition system that mediates the special processing mechanisms that have been attributed to faces. In support of the expertise hypothesis, many face-specific effects have been observed for non-face objects among experts (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Gauthier, Williams, Tarr & Tanaka, 1998). Moreover, it has been shown that non-face objects for which expertise has been acquired can produce activation in a region of the human fusiform gyrus that is

[1] The terms holistic and configural are often used interchangeably in the literature. Configural usually refers to the spatial relations between face parts, and holistic to a tendency to process all of the information present in a face (Gauthier et al., 1998). The distinction between the two concepts is, however, blurred since changes in the configuration of facial features is bound to affect holistic information (see review by Farah, Wilson, Drain & Tanaka, 1998).

preferentially activated by faces (Gauthier, Tarr, Anderson, Skudlarski & Gore, 1999; Gauthier, Skudlarski, Gore & Anderson, 2000).

Assuming that faces exemplify expert recognition, one can use faces to investigate whether or not expertise in visual recognition shares the same characteristics as expertise in other domains. One aspect of expertise that is of particular interest to us is automaticity. Indeed, performing tasks that have become automatic because of extensive training is believed to require little or no attention (e.g., Kristofferson, 1972; Norman & Bobrow, 1975; Schneider & Shiffrin, 1977; Shiffrin, 1977). Our goal was to examine the relationship between attention and the holistic encoding processes that underlie face recognition.

Two landmark studies have provided support for the notion that face recognition relies more heavily on holistic than featural information, with the reverse holding for object recognition. First, Tanaka and Farah (1993) have shown that face parts presented in the context of a studied face are more accurately recognized than when presented in isolation. These results suggest that facial features are not represented in their own but rather as part of a unitary representation of the face as a whole. Scrambled faces, inverted faces, and houses do not show this whole/part advantage, suggesting that these types of stimuli are instead recognized on the basis of featural information. The finding that a whole/part advantage applies to upright but not inverted faces has been replicated (de Gelder & Rouw, 2000; Tanaka, Kay, Grinnell, Stansfield & Szechter, 1998). Second, a study by Tanaka and Sengco (1997) indicated that changing the distance between the eyes disrupted recognition of both faces as a whole and of other unaltered features. Information about the features of a face as well as their relationship thus appear to be combined together to form a holistic representation. As a result, changes in the holistic representation of a face can affect subsequent recognition of its features. Inverted faces and houses did not produce the same results, suggesting that their features and the configural relationship between them are represented independently and therefore do not form a cohesive whole.

The influence of attention on holistic face encoding has been examined by Reinitz, Morrissey and Demb (1994). Using an old/new recognition paradigm, participants were required to either count (divided attention) or ignore (full attention) dots presented across face drawings during the study phase. Recognition was then measured using four types of test stimuli: intact target faces, conjunction faces that were constructed by combining the features of two different target faces, feature faces that were constructed by combining features from a target face and a new face, and completely new faces. Faces studied in the divided attention condition were less accurately recognized as "old" than in the full attention condition. While more "old" responses were given for old faces than for conjunction faces in the full attention condition, the number of "old" responses was equal for the old faces and conjunction faces in the divided attention condition. These results suggest that attended faces were holistically encoded so that their representations could not be matched to the holistic representation generated by the conjunction faces, despite the presence of the old features. In contrast, only the features of ignored faces appear to have been represented in memory since recognition of old and conjunction faces was equivalent. Reinitz et al. (1994) concluded that separate representations exist for featural and configural information and that the role of attention is to link the two together (although see Reinitz, Bartlett, and Searcy, 1997).

Two issues cast doubt on the generality of this conclusion. First, the notion that features and their configuration are represented and stored independently is contrary to prior studies that suggested featural and configural information to be interdependent (Carey & Diamond, 1994; Tanaka & Farah, 1994; Tanaka & Sengco, 1997; Young et al., 1987). Secondly, and perhaps more importantly, the artificial line-drawn faces used by Reinitz et al. (1994) may not have engaged holistic encoding processes to the same extent as photographic images. Indeed, Leder (1996) has shown that manipulations that disrupt holistic information processing have more impact on recognition of photographs than line-drawn faces. As such, line-drawn faces may not be ideal for probing holistic mechanisms. We believe that the question of whether or not attention is required for holistic face processing remains unanswered. In the present study, we have investigated the influence of attention on the composite effect to probe holistic face encoding using face photographs.

The composite effect provides compelling evidence that holistic information processing mechanisms are engaged during face recognition (Young, Hallewell and Hay, 1987; Carey & Diamond, 1994). Composite stimuli are created by aligning the top-half of one face with the bottom-half of another. When they are aligned, the two half faces appear fused together to produce a novel face. Because the holistic representation formed by this new face does not match the stored representations from prior viewing of each face, recognition of the composite becomes difficult. What is most striking about this effect is that the mismatch between stored and perceived holistic representations prevents recognition of the features that make up each half face. However, recognition of the two halves is significantly better if they are horizontally misaligned and therefore do not create a new holistic representation. The difficulty in recognizing upright aligned composites can be attributed to a disruption in both configural encoding, where a new configuration is created when the top half fuses with the bottom half, and holistic encoding, where the task requires recognition of the top part of a face in the context of a whole face (Carey & Diamond, 1994). Stimuli composed of inverted faces can be recognized whether they are misaligned or aligned composites, suggesting that their representations are not holistic. This finding has been replicated using famous as well as unfamiliar faces and throughout different stages of human development (Carey & Diamond, 1994). The composite effect is particularly suited for assessing expert face recognition since the effect holds for non-face objects once expertise in recognizing these objects at the subordinate level is developed (Gauthier et al., 1998).

In our experiments, the composite effect was measured for faces that were encoded in either a full attention or a divided attention condition. Recognition performance was measured for both aligned and misaligned composites that were created using two previously attended or two unattended faces. If the holistic information present in a face can be extracted with little or no attention, then recognition of misaligned composites should be superior to recognition of aligned composites for both attended and unattended faces. Alternatively, if attention is essential for holistic face encoding, then only attended faces should be subject to the composite effect and there should be no difference between recognition of aligned and misaligned composites with unattended faces. As we show in this paper, the composite effect occurs in both full attention and divided attention conditions, suggesting that face processing is holistic with or without the influence of attention.

I.3 Experiment 1

Neisser and Becklen (1975) have shown that overlapped transparent images are a valid tool for the study of selective attention. In Experiment 1, stimuli made up of a face overlapped on a house with 50% transparency were used. During encoding, attention was manipulated toward or away from each face by asking participants to manually delineate the contour of either the face or the house. Six different composite face conditions were evaluated during testing. Attended-aligned and attended-misaligned stimuli were created with faces that were delineated during encoding. Unattended-aligned and unattended-misaligned stimuli were created with faces that were shown while the house was delineated. Distractor-aligned and misaligned stimuli were created with completely new faces. Recognition of aligned stimuli was compared to that of misaligned stimuli for attended and unattended faces separately.

I.3.1 Method

Participants

Thirty-five women and 15 men recruited at McGill University participated in this study. Their ages ranged from 17 to 25. All participants had normal or corrected-to-normal vision.

Apparatus

Participants were tested individually using a Macintosh G3/266 computer with a 21" Macintosh color monitor with a refresh screen rate of 75Hz. The screen was calibrated to linearized luminance values using an Optikon Universal Photometer. The screen was filled by a neutral grey background of 18.6 cd/m^2.

Stimuli

Forty-eight digitized photographs of male faces were obtained from a face database at the University of Essex (http://hpl/essex.ac.uk/projects/vision/allfaces/). The original full-color face images were converted to a 256 grey-level format. The face images were 180 X 200 pixels (subtending 6.5 X 7 degrees of arc at a viewing distance of 57 cm). Forty-eight digitized photographs of houses were obtained from various real-estate web sites. The house images were scaled and cropped to fit in to a 180 X 200 pixels window and converted to a 256 grey-level format. To create the learning stimuli, each face was randomly paired with a house and overlapped with 50% transparency using Abode Photoshop 5.0 software. Learning stimuli were presented on a white 180 X 200 pixels window.

The composite stimuli shown at testing were created by pairing each face with another face on the basis of physical similarity of their contour. All faces were divided into a top and bottom segment by slicing them just below the eyes. Aligned composite stimuli were constructed by positioning the top segment of one face on top of the bottom segment of another face and vice-versa. Misaligned stimuli were created by positioning the nose of the bottom segment of one face next to either the right (Type A) or left ear (Type B) of the top segment of the other face and vice-versa. Composite stimuli were surrounded by a grey window that matched the background.

For each participant, four face pairs (i.e. eight overlapped stimuli) were randomly chosen among the 24 available pairs and used as targets during learning; four other face pairs were randomly chosen and used as distractors for testing.

Procedure

Learning

This study employed a recognition paradigm (Figure 1). Eight overlapped stimuli were presented individually at the center of the screen for 15 s each. A blank screen was shown for 1 sec after the presentation of each stimulus. Before the presentation of each stimulus, a prompt to manually delineate the house or the face was flashed on the screen for 0.6 s. Participants delineated the house in half of the stimuli and the face in the other half in random order. Participants were instructed to devote all of their attention to what they were delineating so that they could recognize that item at testing. Although house delineation was only used to distract attention away from the face, participants were unaware that they would never be tested for house recognition. After the experiment, participants were asked if they felt confident that they were able to focus their attention on the image to be delineated. No data had to be discarded because of a participant's inability to attend to the delineated image.

Testing

During testing, the following six types of stimuli were shown in random order, with two trials per stimulus type: attended-aligned and attended-misaligned stimuli constructed using faces that had been delineated during learning; unattended-aligned and unattended-misaligned stimuli constructed with faces that were shown while the house was delineated during learning. Distractor-aligned and distractor-misaligned stimuli constructed with faces that had not been shown during learning. Half of the misaligned stimuli were shown as type A and the other half as type B. Participants reported whether or not they had seen the top half of each stimulus as quickly as possible by pressing the appropriate key. A blank screen with a black fixation point was shown for 1 sec between stimulus presentations. Percent correct responses and median reaction times for correct responses were recorded.

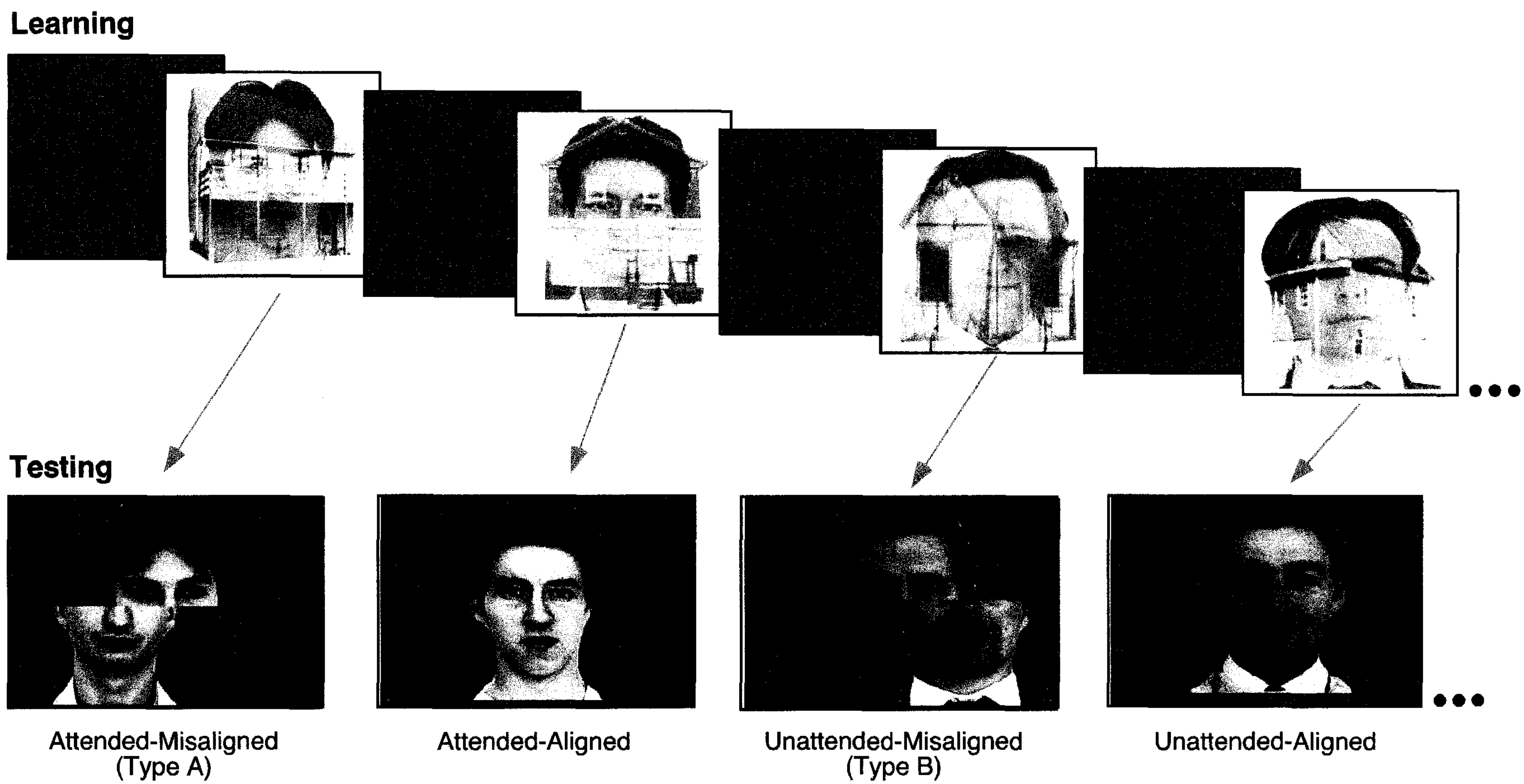

Figure 1. A schematic illustration of the paradigm that was used in Experiment 1. Eight stimuli were shown during learning and twelve during testing. The arrows indicate the corresponding top segment of face stimuli shown during learning. Bottom segments were taken from other faces shown also at learning. Distractor-aligned and distractor-misaligned stimuli are not shown here. The learning and testing stimuli were shown in random order.

I.3.2 Results

Data from one participant was discarded due to a computer error. Percent accuracy and median reaction time data are shown in Table 1. Median RTs are reported here because of the small number of trials per condition. RT data was not considered in our analysis due to the high number of participants who obtained 0% correct responses in one of the conditions tested. *d'* was used as the dependent variable to provide a sensitivity measure which takes into account both target hits and distractor false alarms (Creelman & Macmillan, 1991).

A 2 X 2 repeated-measures ANOVA with Learning Condition (attended and unattended) and Stimulus Type (aligned and misaligned) as variables was performed on the average *d'* measure obtained across participants (Figure 2). The main effects of Learning Condition [$F(1,48)=41.20$, $p<0.01$] and Stimulus Type [$F(1,48)=5.26$, $p=0.03$] were significant. Overall, attended faces were better recognized than unattended ones, and misaligned composites were better recognized than aligned composites. Inspection of the RT data suggests that this difference may have been somewhat amplified by a speed-accuracy tradeoff when recognition of unattended faces was tested. The interaction was not significant [$F(1,48)=0.43$, $p=0.51$].

Planned comparisons were used to evaluate the predictions outlined in the introduction. Recognition of aligned composites was compared to that of misaligned composites for the attended and unattended faces separately. Attended-misaligned stimuli yielded higher *d'* measures than attended-aligned stimuli [$F(1,48)=9.35$, $p<0.01$], as did unattended-misaligned stimuli as compared to unattended-aligned stimuli [$F(1, 48)=4.56$, $p=0.04$]. These results indicate that both attended and unattended faces were subject to the composite effect.

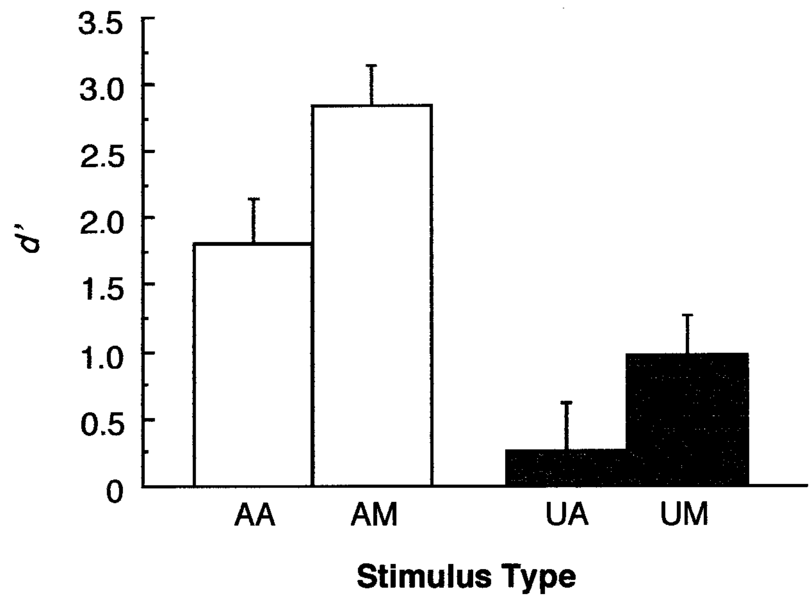

Figure 2. Average d' (49 Ss) obtained in Experiment 1 for the attended-aligned (AA), attended-misaligned (AM), unattended-aligned (UA), and unattended-misaligned (UM) stimuli. Error bars represent ± 1 S.E.

Table 1. Average accuracy and median RT for correctly and incorrectly recognized targets obtained in Experiment 1. Accuracy was calculated by averaging percent correct responses for targets with those obtained for distractors in the corresponding condition. RT data from participants who obtained 0% correct or incorrect responses in one condition were omitted. Standard-deviations are shown in parenthesis.

Test Conditions	Median RT for correct responses	N	Median RT for incorrect responses	N	Average Accuracy	N
Attended Aligned	3.39 (3.13)	41	3.22 (2.40)	29	68.87 (23.68)	49
Attended Misaligned	2.94 (2.62)	46	3.43 (1.98)	23	78.06 (20.17)	49
Unattended Aligned	2.82 (1.73)	22	2.48 (1.42)	43	52.04 (23.84)	49
Unattended Misaligned	2.96 (1.90)	25	2.66 (1.64)	41	60.20 (20.36)	49

I.3.3 Discussion

The results of Experiment 1 indicate that attended faces were, in general, more accurately recognized than unattended faces. This finding provides converging evidence that distracting attention away from a stimulus interferes with its encoding (Tipper, 1985; Tipper & Driver, 1988; Treisman & DeSchepper, 1996; Stankiewicz, Hummel & Cooper, 1998). It is now well established that paying attention to a given stimulus feature is associated with increased activity in the same brain regions that are active during normal viewing of that feature (e.g. Corbetta et al., 1991; O'Craven, Downing & Kanwisher, 1999). Similar attentional modulations have also been observed in the putative face area (Haxby et al., 1994; O'Craven et al., 1999; Wojciulik, Kanwisher & Driver, 1998). The differences in recognition accuracy observed here are likely to reflect increased and decreased activation in neuronal networks recruited during encoding.

The finding that misaligned stimuli were better recognized than aligned ones confirms the composite effect (Young et al., 1987). What we now show is that this effect applies to both attended and unattended faces[2], indicating that attention does not alter the holistic nature of face representations. This result is difficult to reconcile with those of Reinitz et al. (1994). In their study, recognition of conjunction faces constructed by combining the mouth and hair of one learned face with the eyes and nose of another learned face was impaired for full attention but not divided attention conditions.

Three factors may account for the discrepancy between our results and those of Reinitz and colleagues (1994). First, the artificial line-drawn faces they used may not have tapped into the same processes as the more realistic photographed faces that we have used. It may be that holistic encoding can be eliminated by distracting attention away from a face if that face contains little holistic information to begin with (Leder, 1996). Second, it may be argued that our attentional manipulation was not sufficiently effective to eliminate holistic face encoding. Indeed, the long presentation time (15 s) employed here may have allowed participants to pay attention to the faces during delineation of the house. Given that participants were unaware that they would be tested on those faces, and that a main effect of attention was obtained, it is unlikely that this factor was responsible for our finding. Nevertheless, attentional shifts could have occurred in this experiment and therefore we undertook a second experiment (Experiment 2) using an attentional manipulation similar to that employed by Reinitz et al. And finally, composite stimuli (ours) may tap into different perceptual mechanisms than conjunction faces (Reinitz et al). For example, recognition of a part in composites may produce a Stroop-like effect that is less susceptible to attentional manipulation than recognition of a whole face in the conjunction paradigm. Another possibility is that conjunction faces may induce a greater

---

[2] It may be that the attentional manipulation used in our study does not totally eliminate attention to faces and that the use of the word unattended may thus be misleading. As with many other studies on attention, the efficiency of our manipulation was confirmed by the finding of a significantly reduced recognition performance in the attended vs. unattended conditions in both experiments. It is therefore likely that a sufficiently drastic reduction of attention was present to test our hypothesis.

disruption in holistic information than composite stimuli where some of the configural information encoded at learning is preserved in the half face.

To examine if attention distinguishes between configural *versus* holistic information, a control condition was introduced in Experiment 2 where recognition of same-face aligned and misaligned stimuli was tested. In the Gauthier et al. (1998) study, the composite effect was evaluated for novel non-face stimuli called greebles. It was found that same-greeble aligned stimuli are better recognized than same-greeble misaligned stimuli. The authors concluded that when two halves from the same face are misaligned, configural information is disrupted since the relations between the parts are changed. However, holistic information is maintained since the two halves are coming from the same face and no new distracting features are introduced. In contrast, both holistic and configural information would be disrupted in the different-face misaligned composites since the features are displaced (i.e., misaligned condition) and they are shown in a new context because of the presence of the half from the other face (i.e., different-face condition). In Experiment 2, we compared recognition of aligned and misaligned same-face and different-face stimuli to examine if configural and holistic information can be dissociated when face stimuli are used.

I.4 Experiment 2

A different attentional manipulation was used in Experiment 2. During encoding, a target face was presented with letters scrolling quickly from the top to the bottom of the face. The letters formed a sequence of four words. Participants in the divided attention group had to read the letters during the presentation of the face and then report which words had been presented by way of a forced-choice recognition procedure. Participants in the full attention group were told not to read the letters but to use them as cues to move their eyes up and down. At testing, recognition of both target and distractor faces was assessed in four different conditions: same-face aligned, same-face misaligned, different-face aligned, and different-face misaligned.

I.4.1 Methods

Participants

Seventy four students (25 male) recruited at McGill University participated in this study. Their ages ranged from 18 to 26. All participants had normal or correct-to-normal vision.

Apparatus

Same as in Experiment 1 except that the test images were presented on a white background.

Face stimuli

The face stimuli used were the same as in Experiment 1 except that a same-face condition was added. Each face was divided into a top and a bottom segment by slicing it just below the eyes. Same-face misaligned stimuli were created by positioning the nose of the bottom segment to either the right (Type A) or left ear (Type B) of the top segment. For each participant, twelve face pairs were randomly chosen among the 24 available pairs and used as targets during learning; the other face pairs were randomly chosen and used as distractors during testing.

Distractor task stimuli

Thirty-three strings of four words, each word being separated by a period, were generated for the distractor task. For each of the 33 original strings, three matching strings were generated by replacing one or two words from the original string with new words (see Appendix for some examples). Each string produced in this way comprised a total of 26 characters. For each participant tested, each block of four strings was randomly associated with a target face and one of the four strings was randomly chosen as a target for the distractor task.

Procedure

A blocked old/new recognition paradigm was used (Figure 3). Six blocks were used. Each block consisted of one learning stage where four target faces were presented and one testing stage where eight test stimuli were presented.

Learning

For each learning stage, two face pairs were randomly chosen among the target pairs and shown in random order. Each target face was shown along with a distractor task which consisted of the presentation of a sequence of letters appearing at the top and at the bottom of the face. Half of the participants were tested in the full attention condition and the other half were tested in the divided attention condition. Participants were randomly assigned to the two conditions. Participants tested in the full attention condition were instructed to follow the letters with their eyes while paying attention to the face. Participants tested in the divided attention condition were instructed to focus on the letters so as to remember the words that they formed and to ignore the face.

Each learning trial consisted of the presentation of a white window (200 X 200 pixels) in the center of the screen. The first character of the target string was shown at either the top or bottom of the window (counterbalanced) at a random position along a fixed horizontal line. The position of the following characters alternated from top to bottom. The learning face appeared at the center of the white window after the presentation of five characters and stayed on the screen during the presentation of the next 16 characters. The characters were overlaid on the face. The remaining five characters were presented after the face disappeared. Each character was shown for 350 ms, followed by a screen refresh of 20 ms. The face remained on the screen for a total of 6 s.

Presentation of the last character was followed by a blank grey screen for 1 sec, after which the target string and the three distractor strings associated with it were presented. The four strings were shown in the center of the screen, each

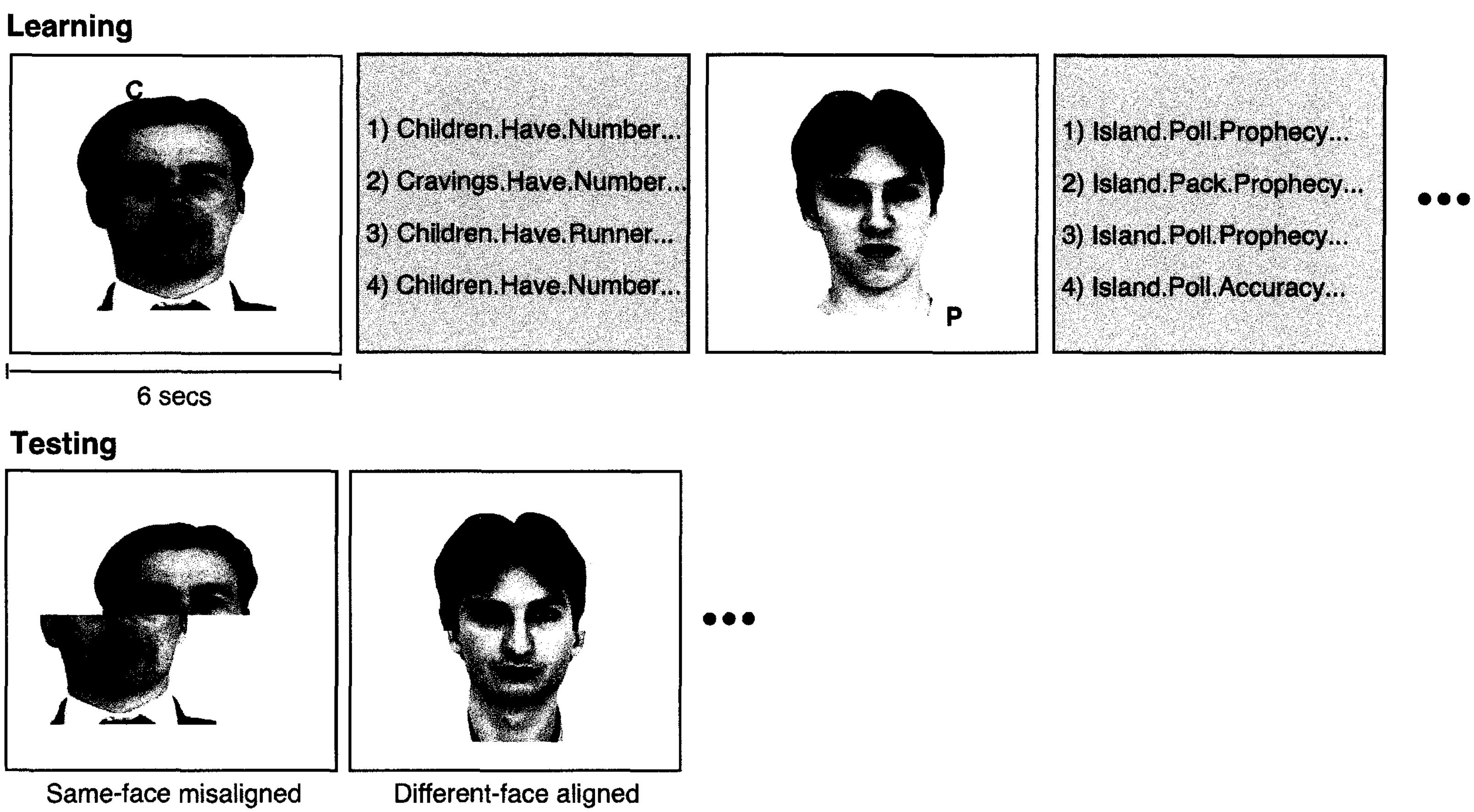

Figure 3. A schematic illustration of the paradigm that was used in Experiment 2. For each block, four stimuli were shown during learning and eight during testing. In the different-face aligned example, the bottom segment was taken from another face shown within the same learning block. The learning and testing stimuli were shown in random order.

string being preceded by numbers one to four. Participants tested in the divided attention condition were instructed to press the key that corresponded to the string shown during learning. Participants tested in the full attention condition were instructed to press the key that corresponded to the string having a third word that differed from the other strings. Negative feedback was provided.

Testing

At testing, eight stimuli were presented, four were used to test recognition of the learned faces and four were distractors. Out of the total 48 test trials shown across testing blocks, six were used to assess each one of the following conditions for both targets and distractors: same-face aligned, same-face misaligned, different-face aligned, different-face misaligned. The two faces from a given pair were tested in the identical condition. For example, if target faces one and two were used to test the different-face misaligned condition, then the two different-face misaligned stimuli created using faces one and two were shown at testing. Presentation order of the different test faces and test conditions was random. Participants reported whether or not they had seen the top half of each stimulus in the previous learning stage as quickly as possible by pressing the appropriate key. A blank screen was shown for 1 sec before the beginning of the next trial. Percent correct responses and reaction times for correct responses were recorded.

I.4.2 Results

Average performance on the distractor task was 80% correct (SD=8.9). Data from four participants was discarded because they obtained less than 2SD (62%) below the mean on the distractor task. Percent accuracy and mean reaction times are shown in Table 2. Recognition performance was assessed using *d'* and RT data.

d'

A 2 X 2 X 2 mixed-design ANOVA with Learning Condition (full and divided attention), Stimulus Type (same- and different-face), and Alignment (aligned and misaligned) as variables was performed on the average *d'* measure obtained across participants (Figure 4). The main effects of Attention [$F(1, 68) = 54.60$; $p < 0.01$] and Stimulus type [$F(1, 68) = 18.57$; $p < 0.01$] were significant. The main effect of Alignment was not significant [$F(1, 68) = 1.63$; $p = 0.20$]. The Stimulus Type X Alignment interaction was significant [$F(1, 68) = 7.18$; $p < 0.01$]. Whereas same-face aligned stimuli yielded slightly higher *d'* values than same-face misaligned stimuli (2.15 and 2.02), different-face aligned stimuli yielded much lower *d'* values than different-face misaligned stimuli (1.31 and 1.83). None of the other interactions were significant.

Table 2. Average accuracy and mean RT for correctly and incorrectly recognized targets obtained in Experiment 2. Accuracy was calculated by averaging percent correct responses for targets with those obtained for distractors in the corresponding condition. RT data from participants who obtained 0% correct or incorrect responses in one condition were omitted. FA—full attention, DA—divided attention. Standard-deviations are shown in parenthesis.

Test Conditions	Mean RT for correct responses	N	Mean RT for incorrect responses	N	Average Accuracy	N
FA Same-Face Aligned	1.12 (0.43)	35	1.77 (0.79)	17	88.57 (9.28)	35
FA Same-Face Misaligned	1.36 (0.47)	35	1.65 (0.53)	20	85 (11.38)	35
FA Different-face Aligned	1.41 (0.66)	34	1.82 (0.88)	34	77.62 (14.68)	35
FA Different-face Misaligned	1.42 (0.66)	35	1.85 (1.10)	26	82.38 (12.75)	35
DA Same-Face Aligned	1.41 (0.9879)	35	1.59 (0.79)	30	68.81 (17.43)	35
DA Same-Face Misaligned	1.43 (0.6560)	34	1.75 (0.90)	30	68.33 (16.01)	35
DA Different-face Aligned	1.69 (1.23)	34	1.63 (0.65)	34	57.14 (14.17)	35
DA Different-face Misaligned	1.50 (0.64)	35	1.58 (0.76)	35	64.29 (14.52)	35

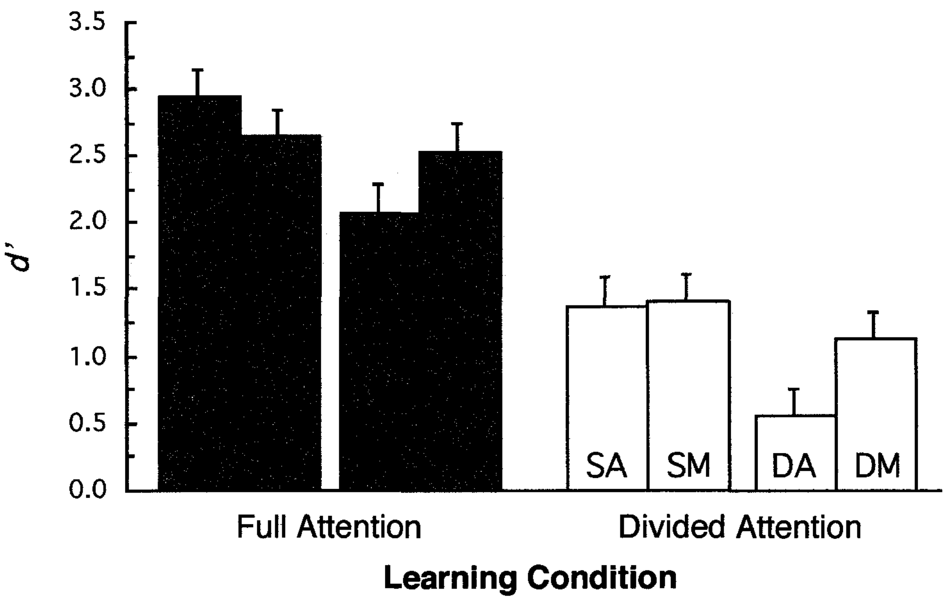

Figure 4. Average *d'* (70 Ss) obtained in Experiment 2 for the same-face aligned (SA), same-face misaligned (SM), different-face aligned (DA), and different-face misaligned (DM) stimuli. Shaded bars represent the full attention condition whereas open bars represent the divided attention condition.
Error bars represent ± 1 S.E.

Planned comparisons were used to test the predictions outlined in the introduction. For the full attention group, there was no difference between recognition of same-face aligned and misaligned stimuli [$F(1, 68) = 1.62$; $p < 0.20$]. Different-face misaligned stimuli yielded higher *d'* values than different-face aligned stimuli [$F(1, 68) = 4.01$; $p = 0.05$]. Results were similar for the divided attention group. While there was no difference between recognition of same-face aligned and misaligned stimuli [$F(1, 68) = 0.02$; $p < 0.88$], different-face misaligned stimuli yielded higher *d'* values than different-face aligned stimuli [$F(1, 68) = 9.11$; $p < 0.01$].

RT

The same analyses as for *d'* were performed for the mean RT for correct responses. Data from participants who obtained 0% correct in one of the conditions tested was discarded for this analysis. The main effects of Attention [$F(1, 65) = 1.81$; $p = 0.18$] and Alignment [$F(1, 65) = 0$] were not significant. The main effect of Stimulus Type was significant [$F(1, 65) = 6.60$; $p = 0.01$]. The Stimulus Type X Alignment interaction was significant [$F(1, 65) = 5.27$; $p = 0.03$]. Consistent with the *d'* data, RT for same-face aligned stimuli were faster than for same-face misaligned (1.277 and 1.394 s), whereas different-face aligned stimuli yielded slower RT than different-face misaligned stimuli irrespective of attention (1.373 and 1.1.459 s). The Alignment X Attention interaction was marginally significant [$F(1, 65) = 3.44$; $p = 0.07$]. It appears that for the attention group, RT for aligned stimuli were faster than for misaligned stimuli (1.269 and 1.373 s). In contrast, for the divided attention group, RTs were slower for the misaligned than for the aligned stimuli (1.569 and 1.459 s). None of the other interactions were significant.

Planned comparisons revealed that correct recognition of same-face aligned stimuli was faster than that of misaligned stimuli for the full attention group [$F(1, 65) = 4.72$; $p = 0.03$]. RT for different-face aligned and misaligned stimuli did not differ [$F(1, 65) = 0.02$; $p = 0.88$]. For the divided attention group, there was no difference between same-face aligned and misaligned stimuli

[$F(1, 65) = 0$]. Consistent with the *d'* data, correct recognition of different-face misaligned stimuli was faster than that for same-face aligned stimuli [$F(1, 65) = 4.45$; $p = 0.04$].

I.4.3 Discussion

The results of Experiment 2 are consistent with those of Experiment 1. First, it was found that recognition was in general more accurate in the full attention *versus* divided attention conditions. Second, the composite effect (Young et al., 1987) was replicated because different-face misaligned stimuli were better recognized than different-face aligned ones. This finding provides further support for the notion that facial features and their configurations are encoded together into a holistic face representation. Most importantly, our results show that this effect holds for both full attention and divided attention conditions, suggesting that attention plays little role in holistic face encoding. Nonetheless, it was surprising to find that the composite effect was stronger for the divided attention than for the full attention group. A ceiling effect may be responsible for this finding. Indeed, thirty-three participants in the full attention group obtained 100% in at least two of the eight conditions tested but only twelve participants in the divided attention group showed such performance.

In Experiment 2, we evaluated whether or not the discrepancy between the results of Experiment 1 and those of Reinitz et al. (1994) could be attributed to a dissociation between configural and holistic encoding. Gauthier et al. (1998) have found that recognition of same-greeble aligned stimuli is faster than that of same-greeble misaligned stimuli. In their opinion, this result indicates a disruption in configural information processing. However, to our knowledge, this effect has never been tested with face stimuli. We found no difference between the same-face aligned and misaligned conditions with regard to *d'* data. However, a difference in RT was obtained for the attention group only. We argue that our RT results do not imply that attention influences configural encoding. Rather, we believe that RTs were more valid in the full than in the divided attention condition because recognition accuracy was superior in that condition. It is generally accepted that RT is a more sensitive measure when the task is relatively easy and accuracy level is high. In difficult tasks, however, accuracy data is more relevant because differences in RT are more easily obscured by variability. Indeed, the RT data was more variable for the divided than for the full attention group. The finding that the Alignment x Attention interaction was almost significant may also be an artifact of overall differences in recognition performance between the two groups.

Another reason why we think that it is unlikely that attention would influence configural but not holistic encoding is that the two processes are closely linked. One would expect that a disruption in configural processing would carry over to holistic processing and therefore minimize the composite effect in the divided attention condition. This was clearly not the case given that the effect was in fact stronger for that group. We therefore conclude that the discrepancy between our results and those of Reinitz et al. (1994) is not due to a dissociation between configural and holistic mechanisms, but rather to their use of line-drawn stimuli that may not probe for holistic face encoding mechanisms (Leder, 1996).

Gauthier et al. (1998) also found that recognition of same-greeble stimuli was more accurate than recognition of different-greeble composites irrespective of alignment. They took this result as evidence for holistic processing in greeble experts. In Experiment 2, the same effect was found whereby same-face stimuli were better recognized than different face stimuli irrespective of alignment for both the full and divided attention groups. This finding provides additional evidence that holistic encoding is not influenced by attention.

I.5 Conclusion

The importance of holistic representation in face recognition has been taken as evidence that faces are analyzed by different mechanisms than objects for which expertise has not been developed. Our results now provide the first evidence that faces can be represented in a holistic fashion irrespective of the allocation of attention during encoding. Using two different attentional manipulations, we have shown that distracting attention away from a face does not eliminate holistic encoding. Our results also suggest that expertise in visual recognition shares the same characteristics as expertise in other domains (e.g., Kristofferson, 1972; Norman & Bobrow, 1975; Schneider & Shiffrin, 1977; Shiffrin, 1977). Holistic face processing, therefore, appears to occur in much the same automatic fashion that is typical of performance with other tasks requiring little or no attention, where automaticity arises due to expertise gained through extensive training. Whether or not other aspects of face encoding can also be performed without attention remains to be determined.

Appendix

Service.Went.Floor.Express
Silence.Went.Floor.Express
Service.Went.Floor.Empress
Service.Went.Frame.Express

Substance.Waged.Voice.Four
Submarine.Waged.Voice.Four
Substance.Waved.Poise.Four
Substance.Waved.Voice.Fire

Night.Crossed.Blue.Turmoil
Night.Crusted.Blue.Turmoil
Fight.Crossed.Blue.Turmoil
Night.Crossed.True.Turmoil

Fever.Take.Painting.Breath
Lever.Take.Painting.Breath
Fever.Take.Stinking.Breath
Fever.Rake.Painting.Breath

Crest.Feel.Canvas.Pedigree
Nests.Feel.Canvas.Pedigree
Crest.Peel.Canvas.Particle
Crest.Feel.Cancer.Pedigree

Weak.Think.Religion.Father
Meak.Think.Religion.Father
Meak.Think.Relieved.Father
Meak.Drink.Religion.Father

Thing.Stand.Telegram.Hopes
Thing.Brand.Telegram.Hopes
Sting.Stand.Hologram.Hopes
Thing.Stand.Telegram.Mopes

World.Push.Figure.Property
World.Lush.Figure.Property
World.Push.Figure.Maturity
World.Push.Finger.Maturity

Preface

The manuscripts provided in Chapter 3 explore a new perceptual rivalry phenomenon produced by overlapped faces but not by overlapped inverted faces or houses. This effect is not only interesting because it shows that specific operations are applied to faces and not to other types of stimuli, but also because it indirectly links face perception with fast processing and informational encapsulation. The results presented in Chapter 3 indicate that two overlapped faces cannot be simultaneously perceived in a holistic fashion. The results also suggest that perception of one face as a coherent whole occurs within one second after presentation, and that an alternation from perception of one whole face to the next occurs within two seconds of presentation. Inverted faces and houses did not readily produce a rivalry effect within the same time constraints. These findings indicate that faces engage specialized operations in that they are more readily perceived as Gestalts than other objects. The findings also suggest that grouping facial features takes place more readily and more quickly than grouping features from non-face images. Finally, the finding that two faces cannot be simultaneously perceived as wholes despite participants' knowledge of the presence of two faces in the display, and despite the task-driven motivation to perceive and recognize both faces, provides indirect evidence for informational encapsulation in face perception.

Chapter 3

Manuscript II

Multistability of overlapped face stimuli is dependent upon orientation

Boutet & Chaudhuri (2001). Multistability of overlapped face stimuli is dependent upon orientation. *Perception*, 30, 743-53.

II.1 Abstract

Stimuli composed of two overlapped faces, one rotated 45° clockwise and the other 45° counterclockwise, produce perceptual rivalry whereby both faces cannot be simultaneously perceived. We obtained subjective and quantitative measures of this rivalry effect and examined if it persists with inverted stimuli. Our results show that upright stimuli are multistable, with alternations occurring from one face to the other within 2 s. Inverted stimuli were instead perceived as ambiguous

in half of the trials, indicating weaker perceptual rivalry in that condition. We suggest that overlapped faces produce perceptual rivalry because each face is readily interpreted into a Gestalt, an effect that in turn is dependent upon orientation

II.2 Introduction

We report a novel multistable phenomenon produced by stimuli composed of two overlapped faces, one rotated 45° clockwise and the other 45° counterclockwise (Figure 1). The stimuli produced perceptual rivalry in that both faces could not be simultaneously perceived. Rather, only one of the faces remained perceptually dominant before alternating to the other face. This phenomenon displays similar characteristics as binocular rivalry effects (e.g. Engel, 1956; Lumer, Friston & Rees, 1998; Yu & Blake, 1992) and reversible figures (e.g., Necker cube, Rubin's reversible goblet, young girl-old woman figure, etc., Attneave, 1971) in that 1) at any one moment only one face can be clearly perceived, and 2) the different percepts alternate periodically.

A possible determinant of this multistable effect is the propensity for faces to be perceived as holistic configurations. A number of studies have shown that configural properties are more important than featural properties during both the perceptual (Farah, Wilson, Drain & Tanaka, 1995; Sergent, 1984) and recognition stages of face processing (Tanaka & Sengco, 1997). There is convincing evidence that perception of a face as a holistic configuration is sensitive to alterations in orientation such that inverted faces may instead be perceived as a collection of segregated facial features (Farah et al., 1995a; Kemp, McManus & Pigott, 1990; Rhodes, Brake & Atkinson, 1993; Sergent, 1984; Tanaka & Sengco, 1997; Young, Hellawell & Hay, 1987). Inverting the stimulus thus provides a test for the possibility that perception of a face as a holistic configuration contributes to the rivalry effect that we observed[3].

[3] The terms 'upright' and 'inverted' are used in this paper to refer to overlapped faces that are actually 45° off the cardinal axes.

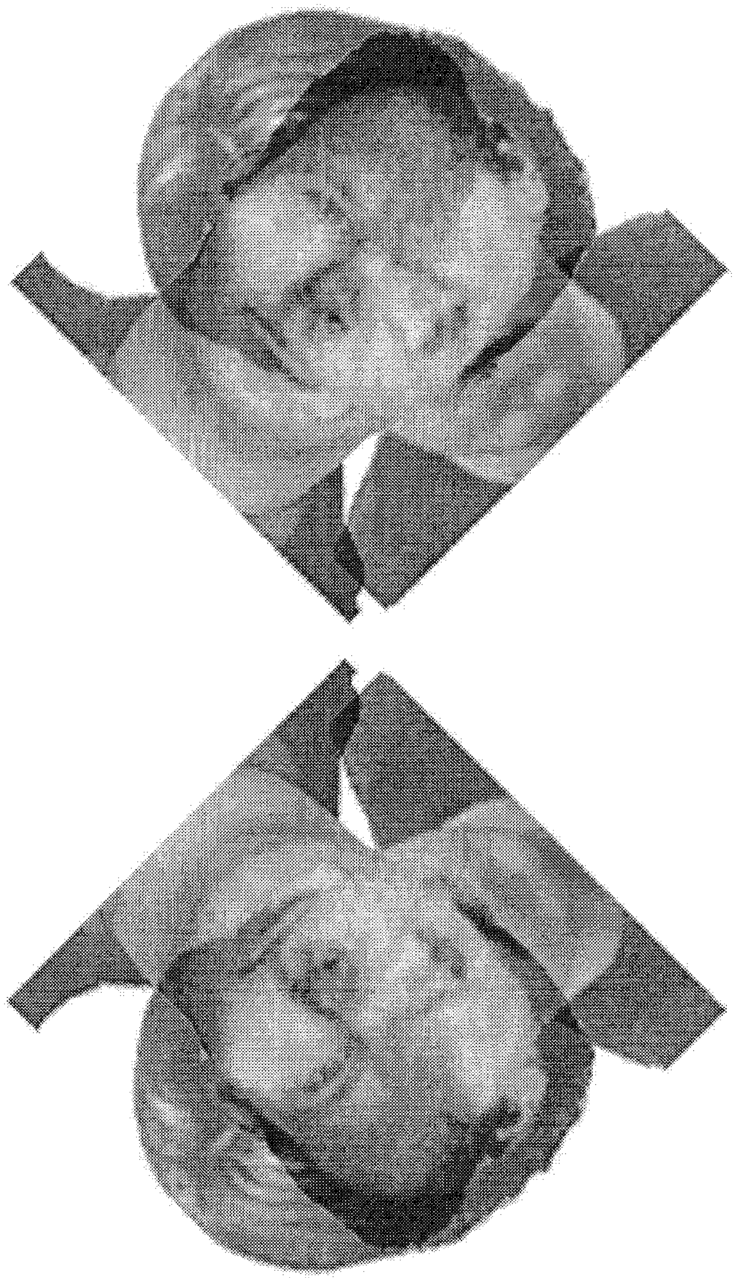

Figure 1. An example of the upright and inverted stimuli used in this experiment.

The experiment reported in this paper served two purposes. First, to provide a measure of the rivalry effect and second, to examine if the effect persists when the orientation is changed. To lend experimental support to the observation that overlapped-face stimuli are multistable, we devised a subjective and quantitative measure of the effect using a sequential matching paradigm. After presentation of each overlapped stimulus, observers were asked to report their subjective experience of the stimulus in a forced-choice task and then to identify both faces among distractors in a recognition task. Overlapped stimuli were presented for one, two, or three seconds.

We reasoned that if the two faces in a stimulus cannot be simultaneously perceived, then short presentation times should allow for the perception of only one of them, while longer presentation times should allow for alternations from one face to the next. Conversely, if there is no rivalry between the two faces, then both should be perceived at all presentation times, or the stimulus could be perceived as ambiguous with neither face being clearly perceived. It is assumed that participants made an attempt to perceive both faces within the presentation time allocated because they knew that recognition of both faces would be tested. As such, perception of only one of the two faces should reflect an inability to perceive both faces simultaneously rather than a tendency to attend to only one of the two faces.

Following this rationale, each trial was classified on the basis of whether one face (*single*), both faces (*double*), or neither face (*no*) had been clearly perceived and accurately recognized. Our prediction was that perceptual rivalry should produce a greater proportion of single than double trials for short presentation times, while the proportion of double trials would be greater than, or equivalent to, that of single trials for longer times because alternations from one face to the next might occur. An absence of rivalry should produce a greater proportion of double than single trials for all presentation times. Finally, it was predicted that ambiguous perception of the stimulus at any given presentation time would result in a greater proportion of no trials than single and double trials.

II.3 Methods

Participants

Twenty-three female and seven male students from McGill University participated in this study. Their ages ranged from 17 to 25. All participants had normal or corrected-to-normal vision.

Stimulus Materials

One hundred twenty digitized photographs of male faces were obtained from a face database at the University of Essex (http://hp1.essex.ac.uk/projects/vision/allfaces/). The original full-color face images were converted to a 256 gray-level format. The face images were 200 x 200 pixels (subtending 7x7 degrees of arc at a viewing distance of 57 cm). All images were unknown to the subjects. They were in full-face view with a homogenous gray-level background. All images were equalized to an average luminance of 47.2 cd/m^2. Half of the faces were tilted clockwise (CW) and the other half counterclockwise (CCW) by 45°. Each CW face was randomly paired with a CCW face and overlapped with 50% transparency using the Adobe Photoshop 5.0 software. Half of the 60 stimuli created were inverted. Half of the observers were tested with these stimuli and the other half were tested with the same CW face-CCW face pairs but with the CW face rotated CCW and vice-versa, and with the inverted stimuli shown upright and vice-versa.

Participants were tested individually using a Macintosh G3/266 computer. The stimuli were presented on a 21" Sony color monitor. The screen was calibrated to linearized luminance values using an Optikon Universal Photometer. A neutral gray background of 18.6 cd/m^2 filled the screen.

Procedure

Participants were instructed that the goal of the experiment was to measure their subjective perception of the overlapped stimuli and to measure their ability to recognize the faces shown during encoding. No suggestion was given as to the possibility that they might experience a rivalry effect. Participants were tested on 30 trials, each consisting of the presentation of one overlapped stimulus, followed by a subjective measure, and then by a quantitative measure of the rivalry effect (see Figure 2). The inter-trial-time was 1s.

Encoding

For each observer, 15 upright and 15 inverted stimuli were randomly chosen and presented during encoding. Each stimulus was presented for either 1, 2, or 3 s. The order of presentation of the stimuli and times were randomized.

Subjective measure—Perceived saliency

The following text was written on the center of the screen 1s after the disappearance of the overlapped stimulus: "Press '1' if you clearly perceived both the CW and CCW images as whole, visible and independent entities. Press '2' if you could only perceive the CW image as a whole, visible and independent entity. Press '3' if you could only perceive the CCW image as a whole, visible and independent entity. Press '4' if the stimulus looked like a scrambled mix of all the features that made up the two images, i.e. if neither image was perceived as a whole, visible and independent entity." Observers were instructed to press the key that best corresponded to their perceptual experience.

Quantitative measure—Recognition accuracy

One second after a key press was recorded, two rows of four images were shown. One row contained CW images and the other CCW images (counterbalanced). The target faces were shown in the same size, orientation (i.e., upright or inverted) and rotation (i.e. CW or CCW) as during encoding. The six distractor faces were randomly chosen from those stimuli that were never shown during encoding. Participants had to press the key corresponding to the image that was shown during encoding for each row in the order of their choice.

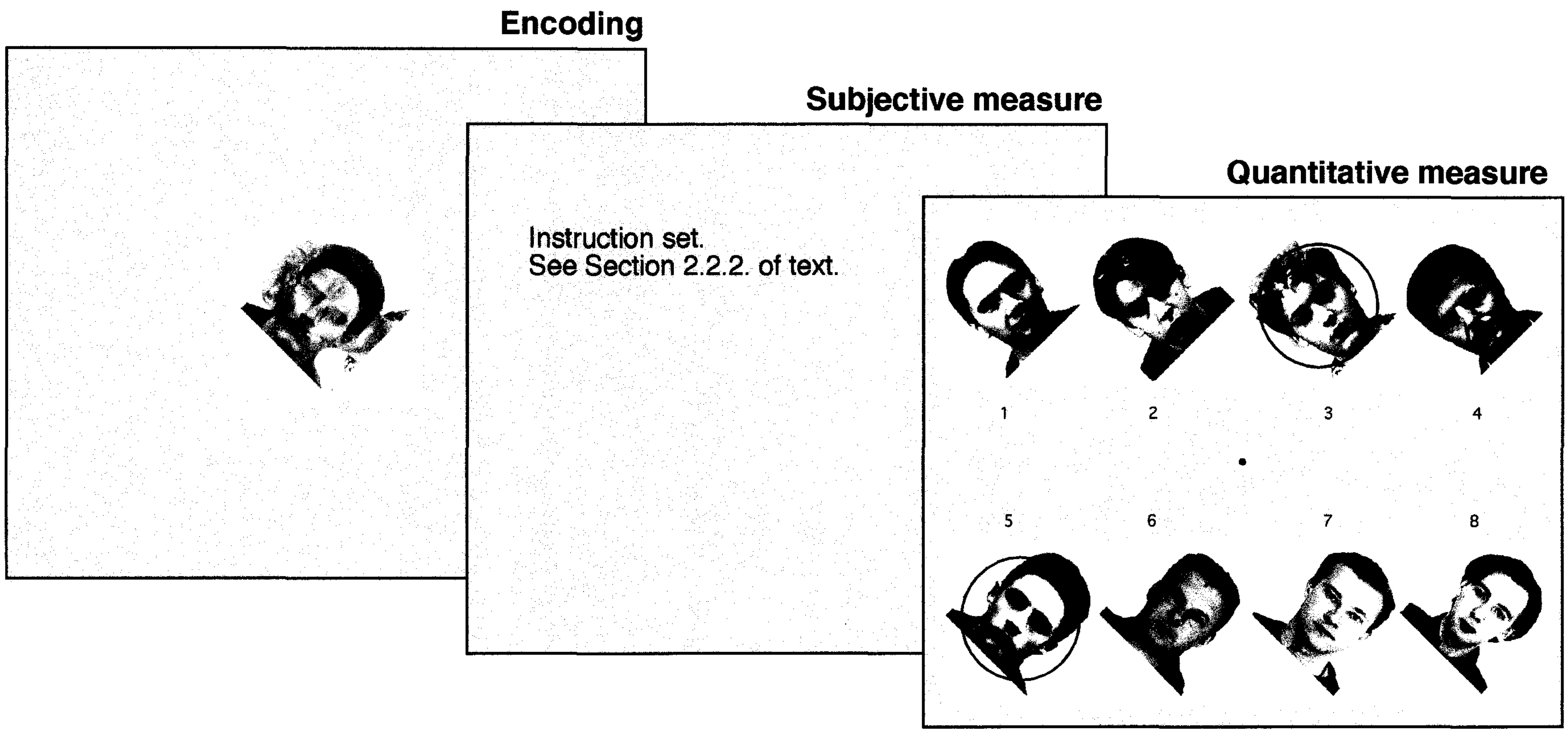

Figure 2. A schematic illustration of the sequential matching procedure that was used in this experiment. Stimuli were presented for 1, 2, or 3 s during the encoding stage. A gray screen was shown for 1 s between the different trial stages and between trials. Perceptual rivalry is evaluated for an upright stimulus in this example.

II.4 Results

Subjective measure

A subjective measure was obtained by calculating the proportion of trials where both faces, a CW face, a CCW face, or neither faces were clearly perceived (Figure 3). Data for two participants were discarded because they reported being able to "clearly perceive both the CW and CCW images as whole, visible and independent entities" on all the trials for both the upright and inverted conditions. However, these participants showed poor recognition performance during quantitative assessment, suggesting that they had difficulties interpreting the descriptions provided.

Perceptual rivalry was evaluated by comparing the proportion of trials where either a single face (single perception–SP) or both faces (double perception–DP) were clearly perceived as indicated by the participant's choice on the subjective measure. Our prediction was that rivalry would produce a greater proportion of SP than DP trials for short presentation times, while the proportion of DP would be greater than, or equivalent to, that of SP trials for longer times because an alternation from one face to the other is more likely. As Figure 3 shows, a short presentation time of upright stimuli produced a greater proportion of SP than DP responses whereas longer presentation times (i.e., 2 and 3 s) showed a marked decrease in SP responses at the advantage of DP responses. This was not the case for inverted stimuli, where the perception of neither face (no perception–NP) prevailed at all presentation times.

A 2 X 3 X 2 repeated-measures ANOVA with Stimulus Orientation (upright and inverted), Presentation Time (1s, 2s, and 3s), and Trial Type (DP and SP) as variables was performed on the average proportion of SP and DP trials obtained across participants. The main effects of Orientation [$F(1, 27) = 59.14$, $p < .01$], Presentation Time [$F(2, 54) = 3.38$, $p = .04$], and Trial Type [$F(1, 27) = 9.26$, $p < .01$] were significant. The Orientation X Trial Type [$F(1, 27) = 4.71$, $p = .04$], Presentation Time X Trial Type [$F(2, 54) = 6.09$, $p < .01$], and Orientation X Presentation Time X Trial Type

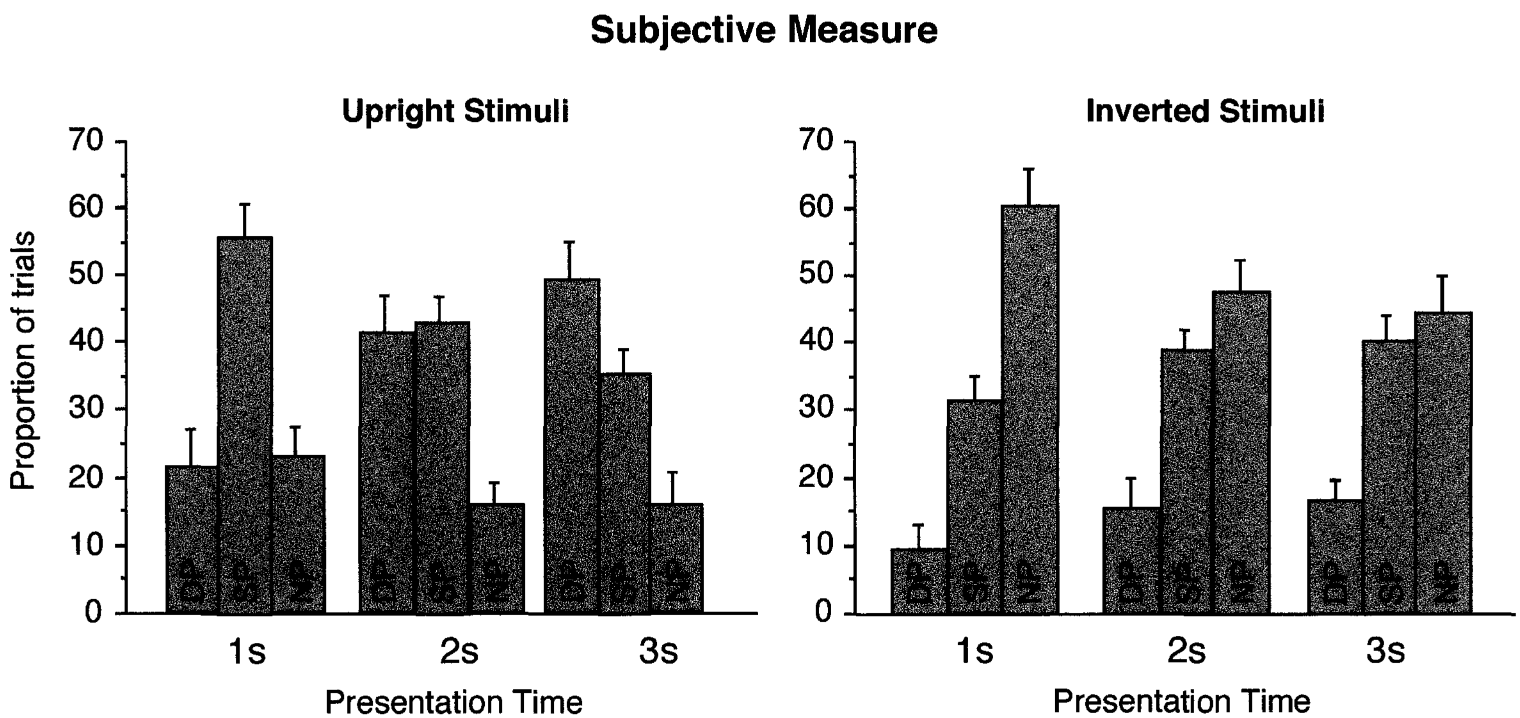

Figure 3. Average proportion of trials obtained for the subjective measure with the upright and inverted stimuli (28 Ss). Trials where both faces (DP), only one face (SP), or neither face (NP) was clearly perceived are illustrated. Error bars represent ± 1 S.E.

[$F(2, 54) = 5.09$, $p < .01$] interactions were also significant. The Orientation X Presentation Time interaction was not significant [$F(2, 54) = 0.30$, $p = .75$].

Planned comparisons were used to evaluate perceptual rivalry for each presentation time and for upright and inverted stimuli separately. Upright stimuli presented for 1s yielded more frequent SP than DP trials [$F(1, 54) = 25.89$, $p < 0.01$], but not stimuli presented for 2s [$F(1, 54) = 0.04$, $p = 0.84$]. Stimuli presented for 3s yielded more frequent DP than SP trials [$F(1, 54) = 1.14$, $p = 0.05$]. These results suggest that upright stimuli presented for 1s induced rivalry in that only one of the two faces was clearly perceived. This finding provides experimental support for the rivalry effect we had casually observed. The difference between SP and DP trials disappeared after 2s, suggesting that longer inspection times allowed alternations from one face to the next to occur. Inverted stimuli presented for 1s [$F(1, 54) = 14.18$, $p < 0.01$], 2s [$F(1, 54) = 10.60$, $p < 0.01$], and 3s [$F(1, 54) = 8.12$, $p < 0.01$] yielded more frequent SP than DP trials. While these results indicate that the inverted stimuli produced perceptual rivalry, further inspection of the subjective reports points to a different conclusion.

We also examined the proportion of trials where participants reported that neither face was clearly perceived (no perception–NP). A 2 X 3 repeated-measures ANOVA with Orientation (upright and inverted) and Presentation Time (1s, 2s, and 3s) as variables was performed on the average proportion of NP trials. The main effects of Orientation [$F(1, 27) = 65.23$, $p < .01$] and Presentation Time [$F(2, 54) = 7.55$, $p < .01$] were significant. The Orientation X Presentation Time interaction was not significant [$F(2, 54) = 0.49$, $p = .61$]. These results indicate that inverted stimuli yielded a greater proportion of NP trials (50%) than the upright stimuli (18%). This finding supports the notion that the inverted and upright stimuli induced a different perceptual experience. It can be appreciated from Figure 1 that while one of the faces immediately pops out in the upright stimulus, the two faces appear to blend together in the inverted stimulus where some effort is required to perceive one of them as a whole configuration. Our interpretation is that segregation of one face from the other was much more difficult with the inverted stimuli than with the upright ones, explaining why neither face was clearly perceived in half the trials in the inverted case.

Quantitative measure

A quantitative measure was obtained by calculating the proportion of trials where both faces, a single face (CW or CCW), or neither face was accurately recognized. As can be seen in Figure 4, single recognition (SR) for upright stimuli is greater than double recognition (DR) at 1s duration whereas at longer times, the difference diminishes (2 s) and actually becomes negligible (3 s). With inverted stimuli, the proportion of SR trials always exceeded the DR trials for all presentation times. More importantly, the number of trials where neither face was accurately recognized (NR) was always greater than for the upright condition.

A 2 X 3 X 2 repeated-measures ANOVA with Stimulus Orientation (upright and inverted), Presentation Time (1s, 2s, and 3s), and Trial Type (DR and SR) as variables was performed on the average proportion of DR and SR trials obtained across observers. The main effects of Orientation [$F(1, 29) = 43.26$, $p < .01$], Presentation Time [$F(2, 58) = 3.28$, $p = .05$], and Trial Type [$F(1, 29) = 41.51$, $p < .01$] were significant. The Orientation X Trial Type interaction was almost significant [$F(1, 29) = 3.05$, $p = .09$]. The Presentation Time X Trial Type interaction was significant [$F(2, 58) = 3.28$, $p = .05$]. The Orientation X Presentation Time [$F(2, 58) = 1.46$, $p = .24$] and Orientation X Presentation Time X Trial Type [$F(2, 58) = 1.75$, $p = .18$] interactions were not significant. These results are not entirely consistent with the subjective measure since the Orientation X Presentation Time X Trial Type interaction was not significant. We discuss this discrepancy in the following section.

Planned comparisons were used to evaluate perceptual rivalry for each presentation time and for upright and inverted stimuli separately. Upright stimuli presented for 1s yielded more frequent SR than DR responses [$F(1, 58) = 12.83$, $p < 0.01$], but not stimuli presented for 2s [$F(1, 58) = 1.30$, $p = 0.26$] and 3s [$F(1, 58) = 0.24$, $p = 0.63$]. In agreement with the subjective data, this finding supports the notion that upright stimuli produced rivalry with alternations occurring within 2s. Inverted stimuli presented for 1s [$F(1, 58) =$

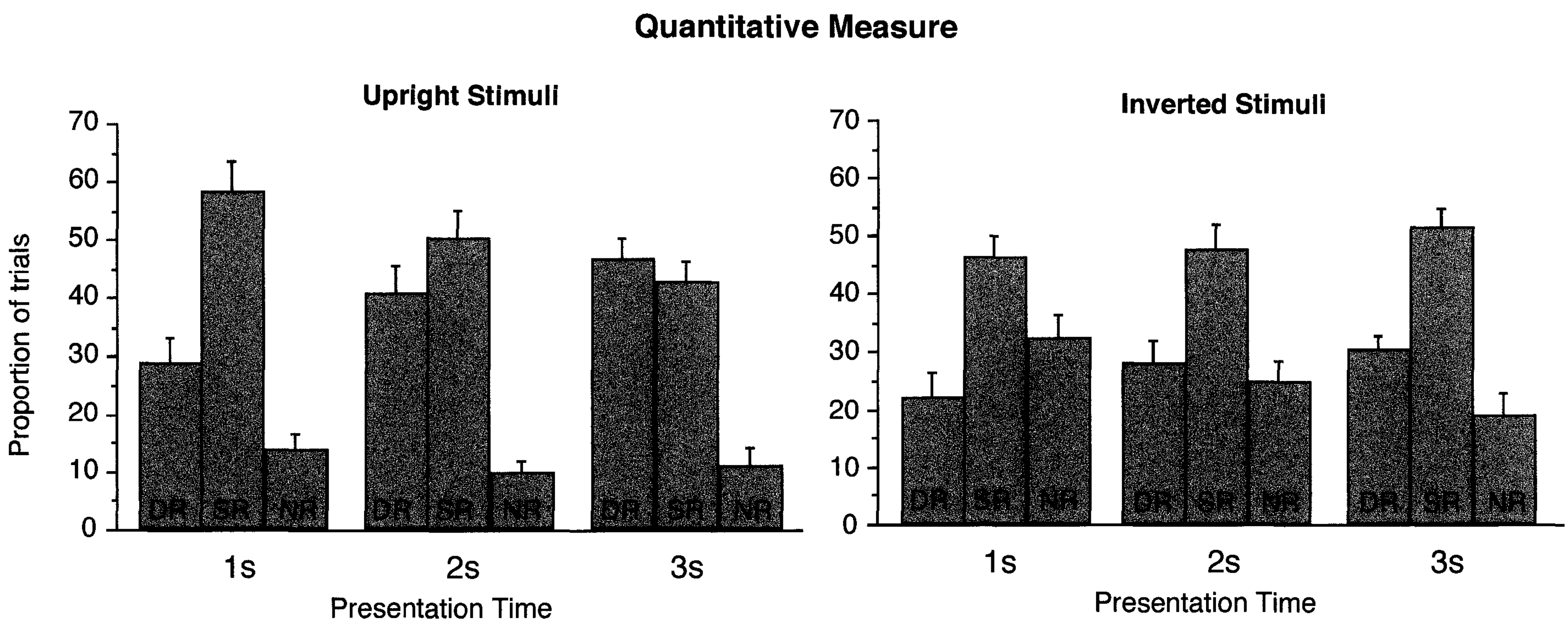

Figure 4. Average proportion of trials obtained for the quantitative measure with the upright and inverted stimuli (30 Ss). Trials where both faces (DR), only one face (SR), or neither face (NR) was accurately recognized are illustrated. Error bars represent ± 1 S.E.

8.59, $p < 0.01$], 2s [$F(1, 58) = 5.57$, $p = 0.02$], and 3s [$F(1, 58) = 6.79$, $p = 0.01$] yielded a greater proportion of SR than DR trials. Since these comparisons support the existence of a perceptual rivalry effect with upright stimuli, the quantitative measure is essentially in agreement with the subjective one.

Because DR trials do not differentiate between an alternation and an absence of rivalry, we also analyzed the proportion of trials where neither face was accurately recognized. A 2 X 3 repeated-measures ANOVA with Orientation (upright and inverted) and Presentation Time (1s, 2s, and 3s) as variables was performed on the average proportion of trials where neither face was accurately recognized (no recognition–NR). The main effects of Orientation [$F(1, 29) = 43.26$, $p < .01$] and Presentation Time [$F(2, 58) = 3.28$, $p = .05$] were significant. The Orientation X Presentation Time interaction was not significant [$F(2, 58) = 1.46$, $p = .24$]. Consistent with the subjective data, these results indicate that inverted stimuli yielded a greater proportion of NR trials (25%) than did the upright stimuli (11%). This finding supports the notion that inverted and upright stimuli induced different perceptual experiences.

Comparison between subjective and quantitative measures

The analyses indicate that the Orientation X Presentation Time X Trial Type interaction was significant for the subjective measure but not for the quantitative one. One possibility for this discrepancy is that a face that was not clearly perceived by the observer could nonetheless have been sufficiently encoded to allow its recognition. To examine this possibility, we compared the subjective and quantitative data relating to those trials where neither face was clearly perceived (NP) or accurately recognized (NR). If faces that are not clearly perceived can nonetheless be accurately recognized, then NP trials should be more frequent than NR trials. A 2 X 2 X 3 repeated-measures ANOVA with Measure Type (NP and NR trials), Orientation (upright and inverted), and Presentation Time (1s, 2s, and 3s) as variables was performed on the average proportion of NP and NR trials. Quantitative data from the two participants whose data were discarded in the subjective measure were not considered for this analysis. The main effect of interest, Measure Type, was significant

[$F(1, 27) = 18.02$, $p < .01$] with NP trials (34%) being more frequent than NR trials (19%). This difference might explain inconsistencies between the two measures whereby participants can accurately recognize a face in a forced-choice procedure, even if they report that the face was not clearly perceived.

Overall recognition accuracy

For the upright stimuli, means (with standard errors in parenthesis) for percent correct recognition for both CW and CCW faces for the 1s, 2s, and 3s presentation times were 57.33 (2.99), 65.00 (3.07) and 68.67 (2.75), respectively. For the inverted stimuli, means were 45.67 (3.56), 52.00 (3.20), and 55.33 (2.79), respectively. The overall effect of inversion on face recognition was evaluated by comparing face recognition for the upright stimuli with that for the inverted stimuli. A 2 x 3 x 2 repeated-measures ANOVA with Orientation (upright and inverted), Presentation Time (1, 2, and 3s), and angle of rotation (CW and CCW) as variables was performed on the recognition accuracy data. The main effect of interest to us, Orientation, was significant [$F(1, 29) = 31.98$, $p < .01$] with an 11% accuracy difference between upright and inverted faces. The main effect of Presentation Time was also significant [$F(12, 58) = 7.73$, $p < .01$]. None of the interactions were significant.

II.5 Discussion

Casual observation of two upright overlapped faces oriented 90° to each other reveals a perceptual rivalry effect where both faces cannot be simultaneously perceived. We obtained subjective and quantitative measures of this phenomenon with both upright and inverted stimuli. We hypothesized that rivalry would lead to the perception and accurate recognition of only one of the two faces for short presentation times while longer times should allow for alternations from one face to the next, and thus in the perception and recognition of both faces. Because trials where both faces are clearly perceived or recognized can reflect either an absence of rivalry or the presence of alternations, we also examined the proportion of trials where neither face was clearly perceived or accurately recognized.

Upright stimuli

Upright stimuli presented for 1 s produced perceptual rivalry as evidenced by the greater proportion of single than double perception and recognition trials. The effect was not apparent at 2 s and 3 s presentation times, suggesting that alternation from one face to the next can occur within 2 s. The visual system is known to segregate one object from another by grouping visual inputs into separate tokens that correspond to known representations. Grouping visual elements into a "Gestalt" can, however, be ambiguous when the tokens that form a stimulus share homogeneous properties, or when alternate groupings are equally plausible. Reversible figures illustrate how the presence of two possible groupings at the same location can produce multistability in that only one grouping is consciously perceived at any given time. Alternation and competition between two plausible groupings has been attributed to mutually exclusive inhibitory mechanisms. According to this model, a particular organization remains dominant until its neural substrate reaches a critical level of satiation or fatigue, at which point the rival organization becomes dominant until it too becomes sufficiently fatigued (Long & Toppino, 1981; Long & Olszweski, 1999; Toppino & Long 1987).

We speculate that one possible explanation for the rivalry effect that we observed may rest on the notion that an upright face is readily represented as a Gestalt (Farah et al., 1995a; Sergent, 1984; Tanaka & Sengco, 1997; Young et al., 1987). As such, faces in our overlapped stimuli may be particularly susceptible to the visual system's propensity to group visual inputs into known representations. Parsing of the visual input into two separable entities at the same location may therefore engage inhibitory processes similar to those implicated in reversible figures. Because each face is readily represented as a holistic configuration, the presence of both plausible groupings at the same location is apparent. However, according to the fatigue/satiation model, passive bottom-up processes would be recruited to suppress one of the two neuronal representations and therefore allow only one of the faces to dominate the perceptual experience.

Inverted stimuli

In our study, inverted stimuli yielded a greater proportion of single than double perception and recognition trials for all the presentation times. While these findings can be taken as evidence that inverted stimuli produced perceptual rivalry, analysis of those trials where neither face was clearly perceived (NP) or accurately recognized (NR) showed that this was not the case. Indeed, inverted stimuli produced a greater proportion of such trials than the upright ones, suggesting the operation of different processes in these two conditions.

One possibility is that faces in the inverted stimuli were not readily encoded as two segregated entities. As a result, the inverted stimuli may have been initially perceived as an ambiguous collection of facial features that did not cohere into a holistic representation. Because the participants were aware that the stimuli contained two faces and that they would be tested on their recognition, we speculate that cognitive "top-down" processes were applied to solve the perceptual puzzle whereby observers "actively constructed" a holistic representation of one face after the other (Pelton, Solley & Brent, 1969). This interpretation is consistent with the notion that learning/decisional mechanisms are involved during the perception of reversible figures (Gregory, 1970; Rock, 1975). According to this view, alternations may be due to a cyclical process of hypothesis testing whereby the perceptual system vacillates between two equally acceptable solutions. This process implies the operation of active mechanisms such as, for example, the allocation of attention to a specific region of a reversible figure during its interpretation (Tsal & Kolbet, 1985). Similarly, prior knowledge of faces may have served to segregate those features that pertained to a given face during perception of inverted stimuli to produce a coherent Gestalt.

Comparison to the classical inversion effect

We have reported a difference of 11% in recognition accuracy between the faces presented in the inverted and upright conditions. This effect size is weaker than previously observed in studies where faces were presented in isolation with a 180° rotation from true upright to inverted (e.g. Diamond & Carey, 1986 ~20%; Yin, 1969 ~ 25%). The fact that the faces in our stimuli were rotated 45° from the cardinal axes, and as such were not truly upright or inverted, may be responsible for the weaker effect. However, the fact that the accuracy and subjective measures yielded statistically significant main effects of Orientation suggests that inversion of our stimuli was sufficient to induce changes in configural information processing similar to those reported in previous studies (Farah et al., 1995a; Kemp et al., 1990; Rhodes et al., 1993; Sergent, 1984; Tanaka & Sengco, 1997; Young et al., 1987).

Possible determinants of the multistable effect

Studies on binocular rivalry indicate that figures whose features can be organized into a coherent figure tend to predominate, and that sensory organization of these figures is often dependent upon orientation. For example, it has been shown that an upright face or "camouflaged Dalmatian" figure dominates over their inverted versions (Engel, 1956; Yu and Blake, 1992). We have shown here that a similar perceptual rivalry effect occurs with facial stimuli whose organization into a Gestalt is particularly sensitive to orientation.

If mechanisms similar to those involved in binocular rivalry are involved in the effect observed here, then our interpretation implies that faces in the upright condition should be organized into a holistic shape before the rivalry can be resolved. It has been shown that competition between binocularly rival complex figures is resolved before information reaches higher visual areas such as the inferior temporal area IT (Sheinberg & Logothetis, 1997; Tong, Nakayama, Vaughan & Kanwisher, 1998). Our interpretation of the multistable face effect suggests that key aspects of configural face information are established in neural loci prior to those where resolution of the rivalry effect takes place.

The tendency for upright faces to be perceived as holistic configurations may be only partially responsible for the multistable effect. Other factors such as familiarity may have also contributed to the effect. Dominance of an upright face or Dalmatian figure over its inverted versions, for example, can be attributed to familiarity (Engel, 1956; Yu & Blake, 1992). In our stimuli, familiarity with upright faces may have facilitated figure-from-ground segregation. However, it is difficult to dissociate familiarity from holistic face encoding because sensory organization of a face as a whole configuration has been attributed to expertise (Carey & Diamond, 1994; Gauthier, Williams, Tarr & Tanaka, 1998; Gauthier & Tarr, 1997; Tanaka & Gauthier, 1997).

Interpretations based on attentional mechanisms have been proposed for other multistable effects (Attneave, 1971; Lumer et al., 1998). For instance, dominance of a given aspect of a reversible figure may be the result of competition between two possible interpretations (Attneave, 1971). We initially speculated that upright stimuli produce perceptual rivalry because encoding of one face as a whole exhausts limited attentional resources. While a study by Reinitz, Morrissey, and Demb (1994) indicates that holistic face encoding is attentionally demanding, we have recently found that perception of faces as holistic configurations requires little attention (Boutet, Gentes-Hawn & Chaudhuri, 2000). The possible role of attention in the multistable phenomenon observed here is thus undetermined.

Attention may contribute to the perception of reversible figures by mediating ocular displacements towards a *focal area* that contains information important for the perception of the predominant interpretation (Garcia-Perez, 1992; Ruggieri & Fernandez, 1994; Scotto, Oliva & Tuccio, 1990; Tsal & Kolbet, 1985). As such, perceptual rivalry could be the result of eye movements towards two different regions of the overlapped stimuli, one region being important for segregation of the CW face, and the other for segregation of the CCW face. While eye movements are likely to be involved in the perception of any reversible figure, they are not the sole determinant of perceptual rivalry because alternations during the perception of reversible figures can occur in the absence of eye movements (Attneave, 1971).

Composite portraiture (Galton, 1883) illustrates one example where overlapped faces do not produce multistability. Composite portraits are created by overlapping upright faces on a photographic film, the result of which is similar to what is obtained by overlapping transparencies in Photoshop software. The fact that non-slanted faces are not multistable and appear to be fused together suggests that rotating faces was a determinant of the rivalry effect that we have reported in this study. Overlapped slanted faces produce a striking difference from the non-slanted condition. With slanted faces, grouping visual features that pertain to each face produces two equally plausible solutions at the same location. This difference may explain why overlapped slanted faces produce rivalry while composite portraits do not.

Concluding remarks

We have suggested that face perception in our upright stimuli is determined by the tendency for faces to be represented as holistic configurations, inhibitory mechanisms are in turn recruited to suppress one representation to allow the alternate to dominate the perceptual experience. Because configural information is not readily constructed from inverted faces, face perception in our inverted stimuli was likely mediated by cognitive processes. Whether or not stimuli created by overlapping other complex objects would also produce rivalry remains to be determined. Given that configural properties are less important during the processing of objects for which expertise has not been developed (Tanaka & Gauthier, 1997; Tanaka & Sengco, 1997), stimuli created by overlapping such objects may be perceived as an ambiguous collection of features and produce a similar effect as that observed with our inverted stimuli. Conversely, previous knowledge of the world could facilitate segregation of two overlapped objects and give rise to a multistable percept. Although we have discussed several possible determinants of the multistable face effect, it is evident that these remain at best speculative and that further study is required to identify the causes of this interesting phenomenon.

Manuscript III

Perceived cohesiveness is an important determinant of the monocular rivalry effect produced by overlapped complex images

Boutet & Chaudhuri. Submitted to *Perception*.

III.1 Abstract

Stimuli composed of two overlapped tilted faces produce perceptual rivalry whereby only one of the faces remains perceptually dominant before alternating to the other face (Boutet & Chaudhuri, 2001). We attributed this monocular rivalry (MR) phenomenon to encoding of faces as a Gestalt because the effect is eliminated by inversion. However, the question of whether perceptual grouping plays a dominant role in MR remains speculative. Here, we report that overlapped houses presented in the same manner as faces failed to show MR but rather were perceived as an ambiguous collection of segregated features, irrespective of orientation. Overlapped negative faces induced an orientation-dependent MR effect similar to that observed with positive faces. Our findings suggest that perceived cohesiveness is an important determinant of the MR effect

produced by overlapped complex images. Other possible factors, such as inhibitory and cognitive processes, are also discussed.

III.2 Introduction

In a previous paper (Boutet & Chaudhuri, 2001), we reported that stimuli composed of two overlapped faces, one rotated 45° clockwise and the other 45° counterclockwise, produce perceptual rivalry whereby only one of the faces remained perceptually dominant before alternating to the other face. Stimuli created by overlapping inverted faces do not produce rivalry but were instead perceived as ambiguous. We suggested that overlapped faces produce rivalry because each face is readily interpreted as a Gestalt, an effect that is in turn dependent upon orientation. In the present paper, we investigate the mechanisms underlying this rivalry effect by examining whether it is dependent on the propensity for each overlapped image to be perceived in a cohesive fashion.

Models of monocular rivalry

Monocular rivalry (MR) effects are similar to those obtained under dichoptic, binocular, conditions (Andrews & Purves, 1997). For example, presentation of two superimposed gratings of opposite orientation and different colors produces perceptual fluctuations in the clarity of one of the two gratings that may lead to a complete extinction of one grating (Campbell & Howell, 1972; Rauschecker, Campbell & Atkinson, 1973). Three general hypotheses have been proposed to account for MR. First, MR may be due to negative afterimages where previous fixation of one orientation produces a 180° out-of-phase afterimage that attenuates the other orientation in the display (Bradley & Schor, 1988; Georgeson & Phillips, 1980; Georgeson, 1984). Although this model can explain MR induced by oriented gratings, the finding that MR also occurs between stimuli which are themselves afterimages suggests that other factors are at play (Crassini & Broerse, 1982).

The second hypothesis asserts that perceptual alternations may arise from a cyclical inhibition of neuronal populations selective for either aspect of a composite stimulus (Campbell & Howell, 1972; Mapperson & Lovegrove, 1978; Rauschecker et al., 1973). The ambiguity produced by the presence of two objects at the same location may be detected in the early stages of visual processing and resolved by inhibition of neural activity associated with the non-dominant image. Alternatively, inhibition may take place because neuronal populations responding to each aspect of a stimulus become 'fatigued' (Atteneave, 1971; Long & Olszweski, 1999). The inhibition model has been adequately applied to some rivalry phenomena, such as that produced by overlapping illusory contours (Fahle & Palme, 1991).

Third, rivalry phenomena may arise from cognitive processes where a decision is made to focus on alternate aspects of unstable stimuli (Anstis, 1975; Gregory, 1970; Walker, 1976). For example, the ambiguity produced by the presence of two equally plausible percepts at the same location may be detected after perceptual grouping and image segmentation operations have taken place. Higher-level selection processes such as attention may then actively select one of the two possible representations (Pelton, Solley & Brent, 1969; Tsal & Kolbet, 1985). Whereas this model is usually mentioned in the context of ambiguous figures, there is evidence that decision-making processes may be involved in MR (Mapperson & Lovegrove, 1984).

These three general models of MR are not necessarily mutually exclusive. Rather, perceptual rivalry effects probably arise from a combination of different factors. The type of mechanism involved may also be dependent on the particular nature of the stimuli employed. Most studies on MR have focused on simple stimuli such as gratings where explanations based on afterimages are well suited for this type of display. However, MR can also be obtained with complex images. We previously reported that overlapping two tilted faces produces rivalry whereby only one face remains perceptually dominant before alternating to the other face (Boutet & Chaudhuri, 2001). Rotating the stimulus by 180° significantly diminished MR and inverted stimuli were instead perceived as an ambiguous mix of facial features (see Figure 1A). We attributed the MR effect

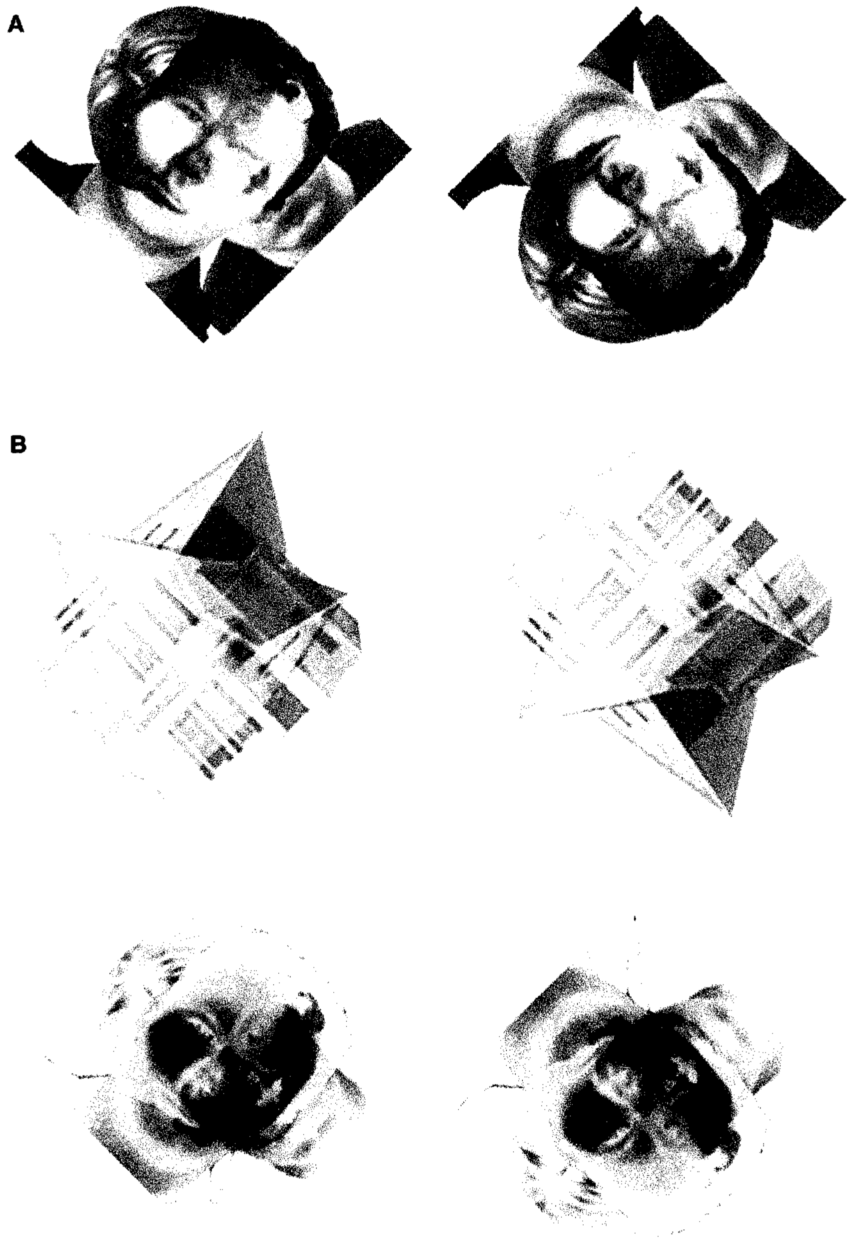

Figure 1. **A** An example of the upright and inverted overlapped face stimuli used in our original experiment (Boutet & Chaudhuri, 2001). **B** An example of the overlapped house stimuli and overlapped contrast reversed face stimuli used in the present experiment.

produced by upright faces to their propensity for being processed in a Gestalt fashion and the current belief that this process is disrupted by inversion (Bartlett & Searcy, 1993; Farah, Wilson, Drain & Tanaka, 1995; Lewis & Johnston, 1997; Sergent, 1984; Tanaka & Farah, 1993; Tanaka & Sengco, 1997, Young, Hallewall & Hay, 1987). Accordingly, it has been shown that perceptual grouping principles play a key role in binocular rivalry (Alais & Blake, 1999; Bonneh, Sagi & Karni, 2001; Engel, 1956; Kovacs, Papathomas, Yang & Feher, 1996; Yu & Blake, 1992). However, the question of whether perceptual grouping plays a similar dominant role in MR remains speculative, largely because simple stimuli were used in past studies, such as lines and gratings. We have used overlapped houses and contrast-reversed faces to address this question in the current study (Figures 1B and 1C).

Holistic versus featural processing of faces and non-face objects

The motivation to use house stimuli rests on the notion that Gestalt processing is more readily applied to faces than to other complex objects, which may instead be encoded on the basis of their features (Farah et al., 1995a; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Likewise, the face-inversion effect, where inverted faces are more difficult to recognize than inverted objects, has been attributed to a greater difficulty in accessing holistic or configural information as compared to featural information after an image has been turned upside-down (Bartlett & Searcy, 1993; Farah et al., 1995a; Leder & Bruce, 2000; Rhodes, Brake & Atkinson, 1993; Yin, 1969). Houses have been used as comparison stimuli in a number of studies to illustrate this distinction (Yin, 1969; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). For example, Tanaka and Sengco (1997) found that changing the distance between the eyes disrupted recognition of both faces as a whole and of other unaltered face features. These results suggest that face encoding is *holistic*, meaning that it involves relatively little part decomposition. In that respect, holistic encoding is comparable to the notion of Gestalt. Tanaka and Sengco also found that houses and inverted faces did not produce the same effect, suggesting that their features and the configural relationship between them are represented independently and therefore do not form a cohesive whole. This type of evidence supports the notion that faces and non-face objects for which expertise has not been developed[4] are analyzed using information at either end of a holistic–featural continuum (Diamond & Carey, 1986; Farah et al., 1995a; Rhodes et al., 1993; Sergent, 1984; Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Young et al., 1987, see Farah, Wilson, Drain & Tanaka, 1998, for a review of the different definitions for these terms). Another motivation for using houses is that they are comparable to faces in many respects. Both faces and houses are regularly encountered in our everyday lives and both share a set of common features that are arranged in a specific manner that renders each exemplar unique.

Holistic versus configural encoding of positive and negative face stimuli

The motivation to use contrast-reversed (or negative) faces is twofold. First, negative faces represent a good control for positive faces because they are readily perceived as a face and because spatial frequency information is preserved after negation. Second, there is some evidence that contrast reversal and inversion tap into different processes that may themselves have a different influence on MR. Contrast reversal is similar to inversion in that it impairs face recognition to a greater degree than object recognition (Bruce & Langton, 1994; Galper, 1970; Hayes, Morrone & Burr, 1986; Liu & Chaudhuri, 1997; Phillips, 1972 for faces; Subramaniam & Biederman, 1997 for objects). Earlier reports suggest that processing *configural* information i.e., the spatial relationship between facial features, is impaired by negation (Kemp, McManus & Pigott, 1990; Lewis & Johnston, 1997). However, a recent study by Hole et al. (Hole, George & Dunsmore, 1999) suggests that *holistic* information i.e., the tendency for faces to be encoded as Gestalts, is not affected by negation. Negation therefore appears to have a specific influence on face processing in that it affects configural but not holistic encoding. Similarly, there is evidence that the face-inversion effect may be due to a specific disruption in configural encoding, while sparing holistic

[4] There is now convincing evidence that behavioral and anatomical differences between face and object processing may not only reflect differences inherent to their visual characteristics but also differences in expertise (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Gauthier, Williams, Tarr & Tanaka, 1998; Gauthier, Skudlarksi, Gore & Anderson, 2000). In this paper, the term object is always used to refer to cases where a discrimination or recognition task is performed on different exemplars of a given non-face object category for which observers are not experts.

information processing (Leder & Bruce, 2000, but see Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Photographic negative face stimuli were thus used to examine whether inversion and negation similarly influence the MR effect produced by overlapped faces. This should in turn clarify the role that configural and holistic encoding processes play in this effect.

Procedure employed to evaluate rivalry

The original evidence for perceptual rivalry with overlapped faces was obtained using a sequential matching paradigm (Boutet & Chaudhuri, 2001). During learning, one overlapped stimulus was shown for one, two, or three seconds. This was followed by a forced-choice procedure to assess subjective perception and by a recognition task where each of the two tilted images shown at learning had to be selected among distractors. We assessed perceptual rivalry using this procedure rather than more traditional measures, such as alternation rates, because we wanted participants to remain naïve with regard to our predictions.

We reasoned that if two images compete for visual awareness, then only one of them should be clearly perceived and hence accurately recognized with short presentation times. Longer presentation times should however, allow for alternations from one image to the next and thus for the perception and recognition of both images. We therefore measured rivalry as a function of the number of trials that fell under two different categories—perception and recognition of only one of the overlapped images (single) or of both images (double). Comparison of these two response types allowed us to evaluate the presence or absence of rivalry, which was operationally defined as a greater proportion of single than double trials at short presentation times, together with a greater proportion of double than single trials at longer presentation times. Conversely, if there is no rivalry between the two tilted images, then both should be perceived at all presentation times and thus produce a greater proportion of double than single trials. A final possibility is that the perceived salience of the images is sufficiently degraded in the overlapped displays and therefore that neither of the images are clearly perceived or accurately recognized (no perception/no recognition).

Predictions

Our prediction for overlapped house stimuli is based on the assumption that houses are not encoded as Gestalts, but rather in a piecemeal fashion similar to that applied to inverted faces (Farah et al., 1995a; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). If the rivalry effect produced by overlapping complex images is dependent on perceived cohesiveness, then we predict that overlapped houses would be ambiguously perceived with a majority of the trials falling under the *no* category irrespective of orientation.

Our prediction for the photographic negative face stimuli is based on the assumption that these images engage holistic encoding processes (Hole et al., 1999) but not configural ones (Kemp et al., 1990; Lewis & Johnston, 1997). One possibility is that the MR effect is determined by a propensity for overlapped images to be encoded holistically. If so, then negative faces should produce an MR effect with single responses being greater than double responses for short presentation times, and double responses prevailing at longer presentation times where an alternation from one face to the next may occur. This holistic process may in turn be affected by inversion such that inverted negative face stimuli would be perceived as ambiguous in a majority of trials. The alternate possibility is that the MR effect is determined by configural information processing. If so, then negative stimuli should not produce an MR effect and would therefore be perceived as ambiguous in a majority of trials irrespective of orientation.

III.3 Methods

Participants

Twenty-six female and four male students from McGill University participated in this study. Their ages ranged from 18 to 24. All participants had normal or corrected-to-normal vision.

Stimulus Materials

The same face images as those used in Boutet and Chaudhuri (2001) were used to create the overlapped photographic negative face stimuli except that their contrast was reversed. One hundred twenty digitized photographs of houses were obtained from various real estate internet sites. The house images were converted to a 256 grey-level format, scaled, and cropped to fit into a 200 x 200 pixels window (subtending 7x7 degrees of arc at a viewing distance of 57 cm). All images were unknown to the participants. House images were equalized to an average luminance of 65.25 cd/m^2. Half of the houses were tilted clockwise (CW) and the other half counterclockwise (CCW) by 45°. Each CW house was randomly paired with a CCW house and overlapped with 50% transparency using the Adobe Photoshop 5.0 software. Half of the 60 stimuli created were inverted. Half of the observers were tested with this set of stimuli whereas the other half were tested with a second set created with the same house pairs but with the CW house being rotated CCW and the CCW house being rotated CW. Overlapped house stimuli that were shown upright in the first set were inverted in the second set and vice-versa. The same applies to the photographic negative face stimuli.

Participants were tested individually using a Macintosh G3/266 computer. The stimuli were presented on a 21" Sony color monitor. The screen was calibrated to linearized luminance values using an Optikon Universal Photometer. A neutral gray background of 18.6 cd/m^2 filled the screen.

Procedure

Participants were instructed that the goal of the experiment was to measure their subjective perception of the overlapped stimuli and to measure their ability to recognize the images shown during encoding. Participants were tested on 60 trials, each consisting of the presentation of one overlapped stimulus, followed by a subjective measure, and then by a quantitative measure of the rivalry effect (see Figure 2). The inter-trial-interval was 1 s.

During the experiment, no suggestion was given as to the possibility that a rivalry effect might be experienced. It is assumed that participants made an attempt to perceive both images within the presentation time allocated because they knew that their recognition would be tested. As such, perception and recognition of only one of the two components should reflect an inability to perceive both images simultaneously rather than a tendency to attend to only one of the two faces.

Encoding

For each observer, 15 different targets were shown from each of the following four stimulus categories—upright house, inverted house, upright negative face, and inverted negative face. These were randomly chosen from the pool of available stimuli. Each stimulus was presented for 1, 2, or 3 s. The order of presentation of the stimuli and presentation times was randomized.

Subjective measure—perceived saliency

The following text appeared on the monitor 1s after the disappearance of the overlapped stimulus: "Press '1' if you clearly perceived both the CW and CCW images as whole, visible and independent entities. Press '2' if you could only perceive the CW image as a whole, visible and independent entity. Press '3' if you could only perceive the CCW image as a whole, visible and independent entity. Press '4' if the stimulus looked like a scrambled mix of all the features that made up the two images, i.e. if neither image was perceived as a whole, visible and independent entity." Observers were instructed to press the key that best corresponded to their perceptual experience.

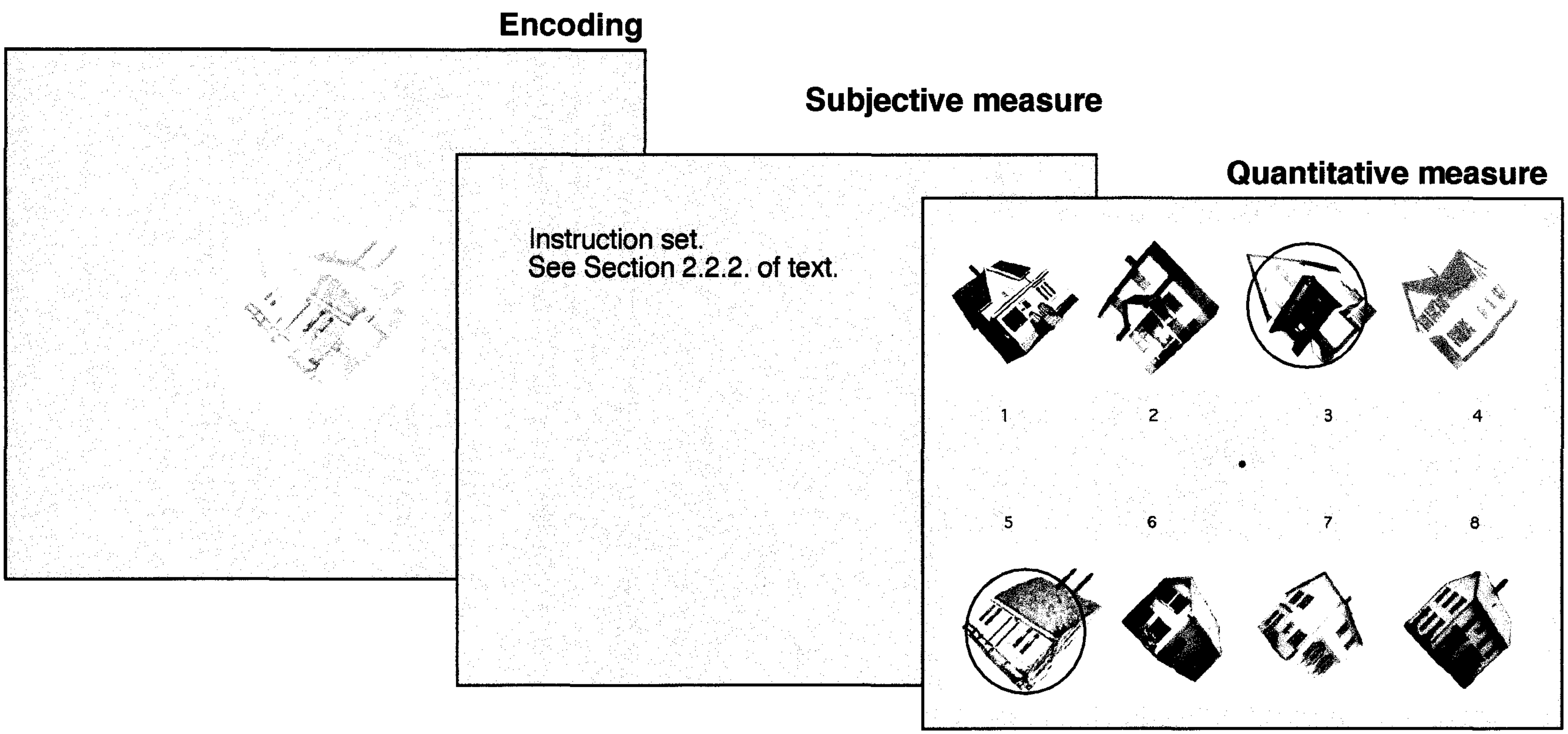

Figure 2. A schematic illustration of the sequential matching procedure that was used in this experiment. Stimuli were presented for 1, 2, or 3 s during the encoding stage. A gray screen was shown for 1 s between the different trial stages and between trials. Perceptual rivalry is evaluated for an upright house stimulus in this example.

Quantitative measure—recognition accuracy

One second after a key press was recorded, two rows of four images were shown. One row contained CW images and the other CCW images (counterbalanced). The target faces were shown in the same size, orientation (i.e., upright or inverted) and rotation (i.e. CW or CCW) as during encoding. The six distractors were randomly chosen from those stimuli that had never been shown during encoding. Participants had to press the key corresponding to the image that was shown during encoding for each row in the order of their choice.

III.4 Results

The presence or absence of perceptual rivalry was evaluated in the same way as in Boutet and Chaudhuri (2001). The proportion of trials where a single image was clearly perceived (single perception–SP) or accurately recognized (single recognition—SR) was compared to that where both images were clearly perceived (double perception–DP) or accurately recognized (double recognition—DR). Our prediction was that rivalry would produce a greater proportion of single than double trials for short presentation times, while the proportion of double trials would be greater than that of single trials for longer times because an alternation from one image to the other may occur. Proportions of trials were compared using a 2 X 3 X 2 repeated-measures ANOVA with Stimulus Orientation (upright and inverted), Presentation Time (1, 2, and 3 s), and Trial Type (double and single)[5]. This analysis was followed by planned comparisons where single and double trials were compared for each presentation time and for upright and inverted stimuli separately. Because we have chosen to evaluate the rivalry effect by comparing single and double trials over time, only the planned comparisons are reported.

[5] Note that this analysis does not take into account the different proportions of chance occurrence for double and single responses. The analysis is meant to highlight differences between double and single responses over time rather than the absolute difference between them.

Overlapped house stimuli

Subjective measure

A subjective measure was obtained by calculating the proportion of trials where both houses (DP), a single CW or CCW house (SP), or neither house (no perception–NP) was clearly perceived as indicated by the participant's choice. As can be seen in Figure 3, SP trials were more frequent than DP trials at all durations for both upright and inverted stimuli. Planned comparisons revealed that upright stimuli presented for 1 s [$F(1, 58) = 23.86$, $p < .01$] , 2 s [$F(1, 58) = 5.55$, $p = .01$], and 3 s [$F(1, 58) = 9.86$, $p < 0.01$] yielded more frequent SP than DP trials. Inverted stimuli presented for 1 s [$F(1, 58) = 20.57$, $p < .01$], 2 s [$F(1, 58) = 31.16$, $p < .01$], and 3 s [$F(1, 58) = 8.45$, $p < .01$] also yielded more frequent SP than DP trials.

These results indicate that SP trials were always more frequent than DP trials regardless of orientation and presentation time. This finding suggests that overlapped houses do not produce perceptual rivalry, and that this effect occurs irrespective of orientation. In addition, this effect is comparable to that previously observed with inverted overlapped face stimuli (Boutet & Chaudhuri, 2001), suggesting that overlapped houses are perceived in a similar fashion as these stimuli.

Ambiguous perception of house stimuli was evaluated using the proportion of trials where neither of the houses was clearly perceived (NP). This was done separately from the analysis applied to DP and SP responses to avoid a total of 100% across cells. A 2 X 3 repeated-measures ANOVA with Orientation (upright and inverted) and Presentation Time (1, 2, and 3 s) as variables was performed on the proportion of NP trials. The main effect of interest, Orientation, was significant [$F(1, 29) = 9.79$, $p < .01$]. These results indicate that there was a difference between upright and inverted stimuli in NP responses. However, both types of stimuli were ambiguously perceived in a large proportion of trials with upright stimuli yielding 40% NP trials and inverted stimuli 50% NP trials.

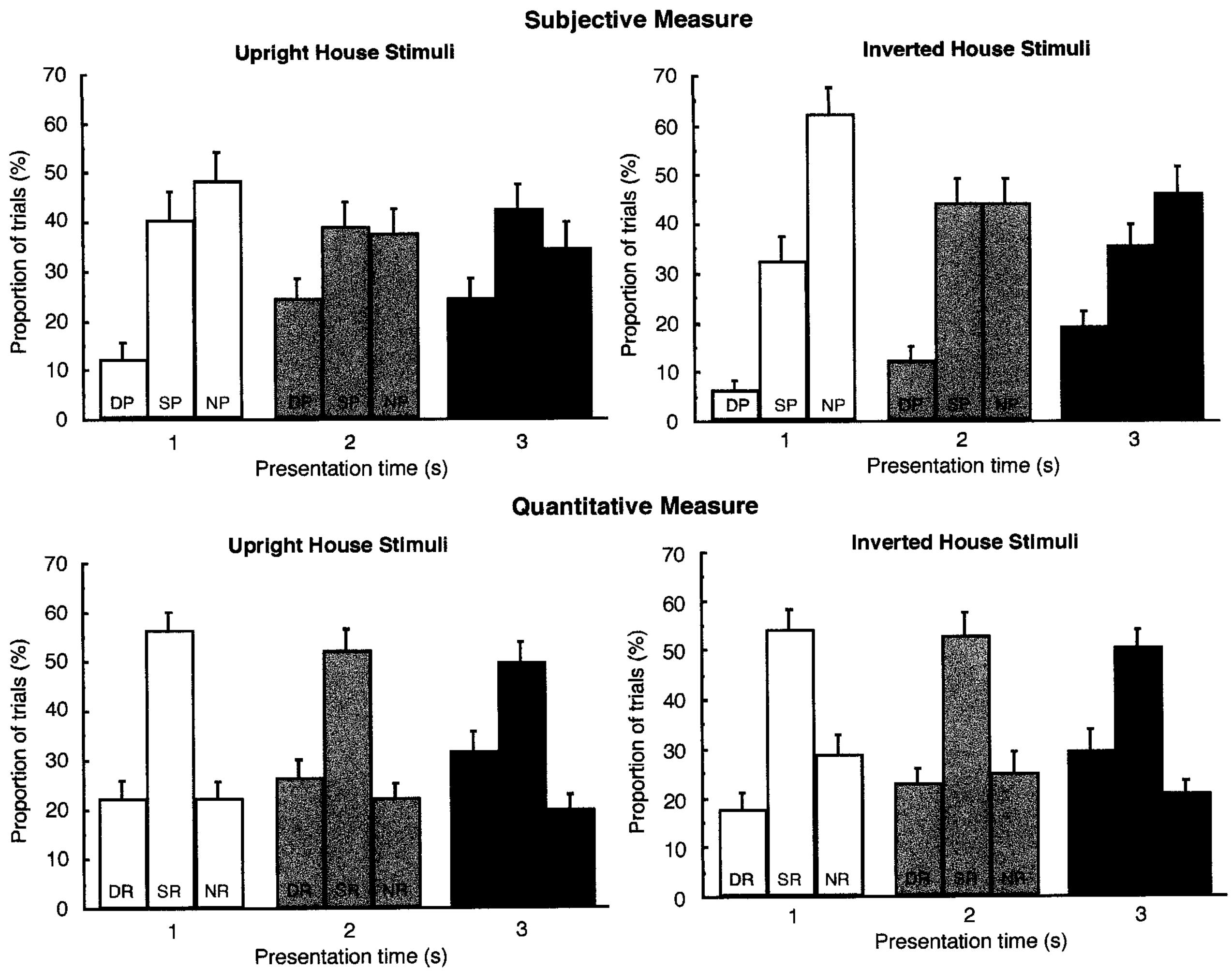

Figure 3. Average proportion of trials obtained for the subjective and quantitative measures with upright and inverted house stimuli (30 Ss). Trials where both houses (DP/DR), only one house (SP/SR), or neither house (NP/NR) was accurately recognized or accurately recognized are illustrated. Error bars represent ± 1 S.E.

Quantitative measure

A quantitative measure was obtained by calculating the proportion of trials where both houses (DR), a single CW or CCW house (SR), or neither house (no recognition–NR) was accurately recognized. As can be seen in Figure 4, SR trials were more frequent than DR trials at all durations for both upright and inverted stimuli. Moreover, the number of NR trials was similar across conditions. Planned comparisons revealed that upright stimuli presented for 1 s [$F(1, 58) = 22.79$, $p < .01$], 2 s [$F(1, 58) = 13.33$, $p < .01$], and 3 s [$F(1, 58) = 6.39$, $p = 0.02$] yielded more frequent SR than DR responses. Inverted stimuli presented for 1 s [$F(1, 58) = 26.51$, $p < .01$], 2 s [$F(1, 58) = 17.75$, $p < .01$], and 3 s [$F(1, 58) = 8.42$, $p < .01$] also yielded a greater proportion of SR than DR trials. These results are in agreement with the subjective data and further suggest that overlapped houses did not produce rivalry irrespective of orientation.

To provide a comparison with the analysis performed on NP responses, the proportion of trials where neither house was accurately recognized (NR) was examined using 2 X 3 repeated-measures ANOVA with Orientation (upright and inverted) and Presentation Time (1, 2, and 3 s) as variables. The main effect of interest, Orientation, was not significant [$F (1, 29) = 2.49$, $p = .13$]. This finding is not consistent with the subjective reports where a main effect of Orientation was obtained with NP responses. This discrepancy is further discussed in section 3.3. below.

Overall recognition accuracy

For the upright stimuli, the means (with standard errors in parenthesis) for percent correct recognition for both CW and CCW houses for the 1 s, 2 s, and 3 s presentation times were 50.00 (3.31), 52.00 (3.34) and 56.00 (3.07), respectively. For the inverted stimuli, the means were 44.33 (2.86), 49.00 (3.24), and 54.33 (3.01), respectively. The overall effect of inversion on house recognition accuracy was evaluated using a 2 X 3 X 2 repeated-measures ANOVA with Orientation (upright and inverted), Presentation Time (1, 2, and 3 s), and angle of rotation (CW and CCW) as variables. The main effect of interest, Orientation, was almost significant [$F(1, 29) = 3.40$, $p = .08$] with a 3% accuracy difference between upright and inverted houses. This finding is consistent with the notion that non-face objects are less sensitive to changes in orientation than faces (Carey & Diamond, 1977; Valentine, 1988; Yin, 1969). Previous studies reported differences of up to 20% between upright and inverted faces (Diamond & Carey, 1986; Yin, 1969) and a difference of 11% was obtained in our original study with face stimuli (Boutet & Chaudhuri, 2001).

Overlapped photographic negative face stimuli

Subjective measure

The same analyses as those performed for the overlapped houses were performed on the data obtained for the overlapped negative faces. Figure 4 shows that perception of upright stimuli was modulated by presentation time, with the proportion of SP trials being greater than that for DP trials at the 1 s presentation times, and DP trials being greater than SP trials for the 3 s presentation. Inverted stimuli yielded a greater proportion of SP than DP trials at all times, with NP trials prevailing at all three durations.

Planned comparisons revealed that upright stimuli presented for 1 s yielded more frequent SP than DP trials [$F(1, 58) = 17.90$, $p < 0.01$], but not for stimuli presented for 2 s [$F(1, 58) = 1.19$, $p = .28$]. Upright stimuli presented for 3 s yielded more frequent DP than SP trials [$F(1, 58) = 915.67$, $p < 0.01$]. Inverted stimuli presented for 1 s [$F(1, 58) = 27.45$, $p < .01$], 2 s [$F(1, 58) = 4.19$, $p = .05$], and 3 s [$F(1, 58) = 4.19$, $p = .05$] yielded more frequent SP than DP trials.

It appears that upright overlapped negative faces produce perceptual rivalry because SP trials were more frequent than DP trials for the 1 s presentation times, and DP trials prevailed at the 3 s presentation time where an alternation from one face to the next would have been possible. For inverted negative faces, SP trials were always more frequent than DP trials, suggesting that these stimuli did not readily produce a rivalry effect. These results are similar to those obtained with stimuli created by overlapping positive faces (Boutet & Chaudhuri, 2001).

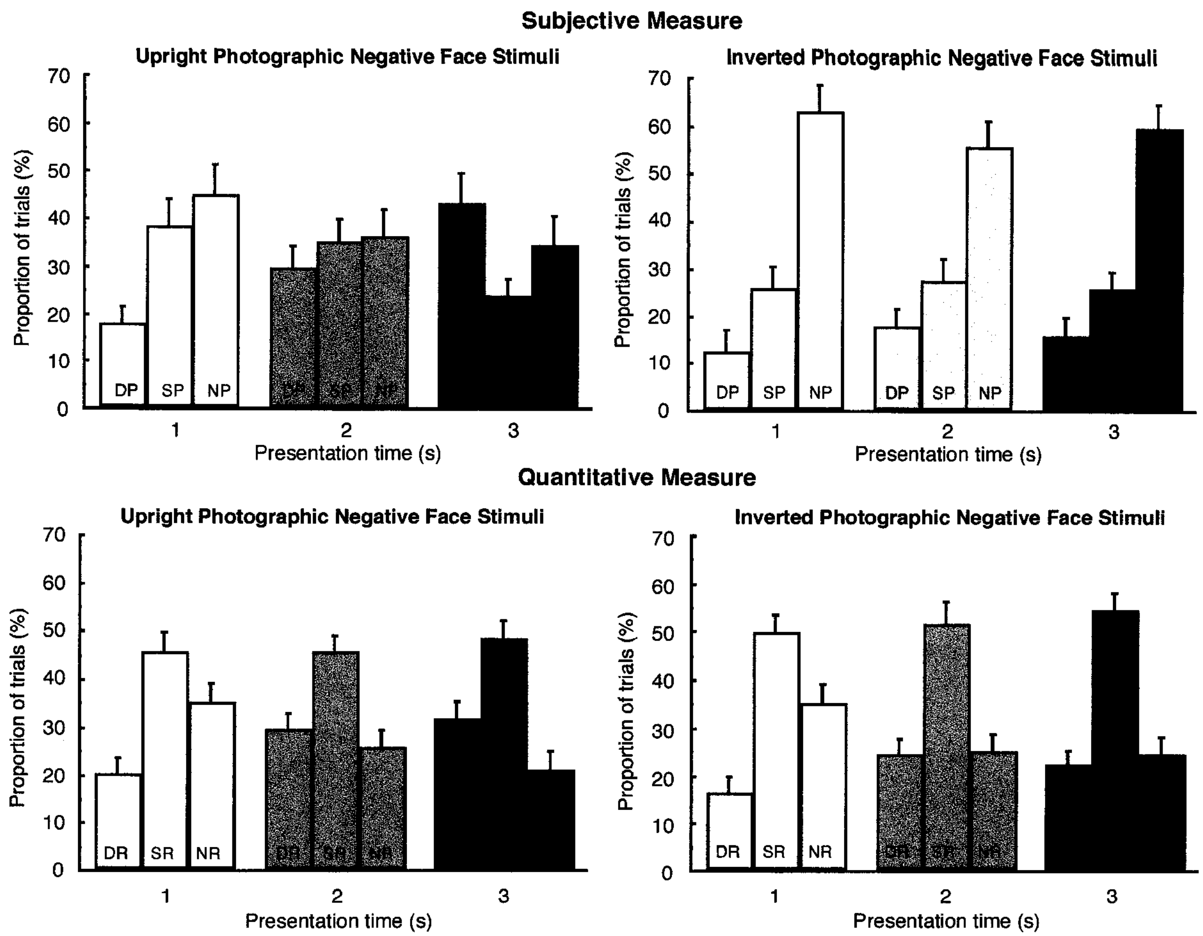

Figure 4. Average proportion of trials obtained for the subjective and quantitative measures with upright and inverted negative face stimuli (30 Ss). Trials where both faces (DP/DR), only one face (SP/SR), or neither face (NP/NR) was clearly perceived or accurately recognized are illustrated. Error bars represent ± 1 S.E.

Upright negative stimuli yielded 40% NP trials and inverted ones 60% NP trials. A 2 X 3 repeated-measures ANOVA with Orientation (upright and inverted) and Presentation Time (1, 2, and 3 s) as variables revealed that this difference was significant [$F(1, 29) = 29.31$, $p < .01$]. This finding suggests that inversion had a greater influence on the perception of negative overlapped faces, which yielded a 20 % orientation effect, than overlapped houses, which yielded a 3% orientation effect.

Quantitative measure

Figure 4 shows that upright and inverted photographic negative face stimuli produced a similar recognition performance with SR trials prevailing for all presentation times. Planned comparisons revealed that upright stimuli presented for 1 s [$F(1, 58) = 22.16$, $p < .01$], 2 s [$F(1, 58) = 8.84$, $p < .01$], and 3 s [$F(1, 58) = 9.59$, $p < 0.01$] yielded more frequent SR than DR trials. Inverted stimuli presented for 1 s [$F(1, 58) = 38.36$, $p < .01$], 2 s [$F(1, 58) = 25.80$, $p < .01$], and 3 s [$F(1, 58) = 35.36$, $p < .01$] also yielded a greater proportion of SR than DR trials.

The subjective and quantitative results obtained for the upright stimuli were not consistent. Whereas the subjective results suggest that upright stimuli produced a perceptual rivalry effect that was dependent upon orientation, the quantitative results suggest that both upright and inverted stimuli failed to produce rivalry with SR trials prevailing irrespective of presentation time and orientation. This discrepancy is further discussed in section 3.3. below.

Upright photographic negative stimuli yielded 27% NR trials and inverted ones yielded 28% NR trials. A 2 X 3 repeated-measures ANOVA with Orientation (upright and inverted) and Presentation Time (1, 2, and 3 s) as variables revealed that this difference was not significant [$F(1, 29) = 0.10$, $p = .75$]. This finding suggests that recognition of negative stimuli was not influenced by inversion.

Overall recognition accuracy

For the upright stimuli, means (with standard errors in parenthesis) for percent correct recognition for both CW and CCW CR faces for the 1 s, 2 s, and 3 s presentation times were 42.67 (3.06), 52.00 (2.90) and 55.33 (3.24), respectively. For the inverted stimuli, means were 40.00 (3.01), 49.00 (3.10), and 49.67 (2.95), respectively. The overall effect of inversion on recognition of photographic negative faces was evaluated by comparing recognition for the upright stimuli with that for the inverted stimuli. A 2 X 3 X 2 repeated-measures ANOVA with Orientation (upright and inverted), Presentation Time (1, 2, and 3 s), and angle of rotation (CW and CCW) as variables was performed on the recognition accuracy data. The main effect of interest to us, Orientation, was not significant [$F(1, 29) = 2.86$, $p = .10$] despite a 4% accuracy difference between upright and inverted faces.

Comparison between the subjective and quantitative measures

It is puzzling that both the analyses performed on the house stimuli data and those performed on the negative face stimuli data produced significant effects of Orientation with the subjective NP responses but not with the quantitative NR responses. In addition, the subjective measure but not the quantitative measure revealed an orientation-dependent rivalry effect for the negative face stimuli. We further examined these discrepancies by calculating whether or not these two measures were consistent for any given trial. The average proportion of trials where the answer on the subjective and quantitative measures matched was calculated for each participant. A 2 X 2 X 3 repeated-measures ANOVA with Stimulus Type (house and face), Orientation (upright and inverted), and Presentation Time (1, 2, and 3 s) was performed on these average match-scores. Only the main effect of time was significant [$F(1, 29) = 16.58$, $p < .01$] with the 1 s presentation yielding the best match (44%) over the 2 s (29%) and 3 s (27%) presentation times.

These results suggest that image type and orientation did not determine whether these two measures corresponded or not. Hence, the discrepancies observed do not appear to arise from potential differences introduced by these variables, but rather from differences inherent to the type of process that each measure taps into. What factors may account for this difference, and which of these two measures provides a more valid assessment of perceptual rivalry with respect to our operational definition, is further considered in the Discussion.

III.5 Discussion

We have previously found that two upright overlapped faces oriented 90° to each other produce a perceptual rivalry effect whereby only one of the faces remains perceptually dominant before alternating to the other face (Boutet & Chaudhuri, 2001). Inverted overlapped faces do not produce this effect but are rather perceived as an ambiguous collection of features. We attributed these results to the propensity for upright faces to be processed as Gestalts and the current belief that this process is disrupted by inversion (Bartlett & Searcy, 1993; Farah et al., 1995a; Lewis & Johnston, 1997; Sergent, 1984; Tanaka & Farah, 1993; Tanaka & Sengco, 1997, Young et al., 1987).

The notion that binocular rivalry effects are mediated by perceptual grouping principles is now well established (Alais & Blake, 1999; Bonneh, Sagi & Karni, 2001; Engel, 1956; Kovacs et al., 1996; Yu & Blake, 1992). However, it remains unclear whether such grouping principles also operate in the monocular rivalry (MR) effect produced by overlapped complex images. To address this issue, we examined whether stimuli that are encoded at different points along the holistic-featural continuum produce an MR effect.

Overlapped house stimuli

We had predicted that overlapped house stimuli would not readily produce a rivalry effect because object encoding relies less heavily on holistic information than does face encoding (Farah et al., 1995a; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). Overlapped houses produced a greater proportion of single perception and recognition (SP/SR) trials than double perception and recognition (DP/DR) ones, irrespective of stimulus duration and orientation. We also found that overlapped houses were perceived as ambiguous in 40% to 50% of the trials.

Together, these results suggest that overlapped houses do not readily produce a monocular rivalry (MR) effect and that they are instead perceived as an ambiguous collection of segregated features.

Overlapped houses produced a pattern of results almost identical to that previously obtained with overlapped inverted faces (Boutet & Chaudhuri, 2001). This result is consistent with the current belief that inverted faces are encoded in a piecemeal fashion similar to that applicable to non-face objects (Leder & Bruce, 2000; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). These data together suggest that the monocular rivalry (MR) effect produced by overlapped complex images is mediated by the perceived cohesiveness of their constituent features.

Overlapped photographic negative face stimuli

The subjective results suggest that upright negative stimuli induced a rivalry effect because SP trials were greater than DP trials for the one second presentation time, with the reverse being true for the three second presentation time. However, inverted negative faces yielded greater SP trials at all time points, suggesting that the rivalry effect was dependent upon orientation. In contrast, the quantitative results showed that SR trials were more frequent than DR trials irrespective of orientation and presentation time. Thus, the subjective and quantitative results were not consistent in that upright faces displayed greater SP than DP trials only at a short presentation time, but SR trials prevailed over DR for all presentation times.

For the subjective measure, participants were asked to report whether both, only one, or neither image was clearly perceived as a whole independent entity, a judgment that probably reflected their ability to segment one image from the other. In contrast, the quantitative measure does not necessarily reflect whether an image is perceived as a whole because recognition on the basis of a single salient feature is possible even if the image was not coherently perceived. Therefore, the discrepancy between the two measures may rest in the fact that the subjective measure reflects immediate perception of figure-from-ground during encoding, while the quantitative measure reflects feature-based recognition processes during testing. Because our aim was to evaluate whether overlapped images produce a rivalry effect whereby one image is readily perceived as 'popping out' of the other background image, we believe that the subjective measure provides a more valid assessment of this effect than the quantitative one. We therefore interpret our findings on the basis of the subjective measure and suggest that negative face images produced a rivalry effect that was dependent upon orientation.

The finding that both the house and the negative face data produced significant effects of Orientation with the subjective *no* responses, but not with the quantitative ones, also needs to be considered. Feature-based recognition is believed to be unaffected by orientation because a salient feature can be easily identified even after inversion (Carey & Diamond, 1994; Rhodes et al., 1993; Yin, 1969). This may explain why the effect of orientation was not significant with the quantitative recognition measure. In contrast, segmentation of the overlapped images is bound to require some prior knowledge of what the figure looks like, especially if neither figure automatically pops out. Lack of familiarity with inverted houses and negative faces may have therefore significantly influenced the perception of the overlapped stimuli and resulted in the significant effect of orientation observed with the *no* subjective responses.

Our finding that the rivalry effect produced by overlapping faces is affected by inversion (Boutet & Chaudhuri, 2001) but not by contrast reversal suggests that these two manipulations tap into different processes that may in turn have a differential influence on MR. One possibility is that inversion disrupts both holistic and configural information whereas contrast reversal disrupts only the latter, as suggested by previous studies. Kemp et al. (1990) and Lewis and Johnston (1997) have shown that manipulations that affect the spatial relations between face parts have a greater effect on recognition of positive than negative faces. In contrast, Hole et al. (1999) have shown that the composite effect, where presentation of the bottom half from one face interferes with recognition of the top half from another face, is equivalent for positive and negative faces. Although this finding appears to suggest that negation disrupts configural information processing, the difficulty in recognizing composites may instead arise from holistic encoding processes (Carey & Diamond, 1994). The finding that the composite effect is not affected by negation, but is eliminated by inversion, further suggests that these two manipulations tap into different processes.

Studies that have employed recognition paradigms have produced conflicting evidence with regard to the influence of inversion on holistic versus configural encoding processes (Leder & Bruce, 2000; Tanaka & Farah, 1993; Tanka & Sengco, 1997). However, perception of face parts in an interactive fashion has been shown to be dependent upon orientation (Sergent, 1984). Hence, the evidence currently available suggests that inversion disrupts perception of faces as Gestalts whereas negation does not. This would explain why the rivalry effect produced by overlapping faces is determined by orientation, irrespective of whether the faces are positive or negative, and supports our conclusion that perceived cohesiveness is an important determinant of rivalry.

Possible determinants of the rivalry effect produced by overlapped complex images

Three general models have been proposed to account for other MR effects—masking due to afterimages, cyclical inhibition of neural systems, and cognitive decisional processes. Afterimages are not responsible for the MR effect obtained by overlapping complex images because each image in the overlapped display does not produce an out-of-phase afterimage of the other.

Our results suggest that perceived cohesiveness is an important determinant of the rivalry effect produced by overlapping complex images. Such rivalry may be the result of inhibition at a stage of visual processing where each image is represented as a coherent whole (Leopold & Logothetis, 1996; Sheinberg & Logothetis, 1997). For example, competition could take place in the temporal cortex, which in monkeys has been shown to contain neurons selective for whole faces (Gross, 1992; Perrett, Hietanen, Oram & Benson, 1992). It may be the case that a network of face-encoding neurons cannot be simultaneously activated by two different faces at the same location because it represents an unnatural and implausible condition. A resolution of this ambiguity may be achieved via inhibition of one of the two representations. Inhibition of neural networks that provide cohesive representations of competing images has previously been proposed for binocular rivalry between complex figures (Tong, Nakayama, Vaughan & Kanwisher, 1998; see also Bonneh et al., 2001; Kovacs et al., 1996; Lumer, Friston & Rees, 1998). We suggest that similar mechanisms may be involved in the perception of overlapped faces.

We believe that cognitive processes are more likely to be recruited during the perception of images that are encoded in a piecemeal fashion, such as houses and inverted faces, than during the perception of images that are readily perceived as Gestalts, such as upright positive and negative faces. Indeed, the speed and ease with which one face pops out of a given overlapped face display suggests that processing may be more automatic in nature. Cognitive top-down processes may instead be operative during perception of overlapped houses and serve to *actively construct* a holistic representation of each image in the stimulus (Pelton et al., 1969), which would otherwise be perceived as an ambiguous collection of segregated features.

It is possible that with sufficient time, top-down processes can be used to segregate one house from another, or one inverted negative face from another, to produce a rivalry effect comparable to that observed with the faces. Similarly, it is possible that the greater proportion of single perception responses observed with these stimuli at a long presentation time is not indicative of an absence of rivalry, but rather indicates a rivalry effect where an alternation from one image to the next takes more than three seconds. Prolonged inspection of the overlapped house and inverted negative face stimuli suggests that each image can be grouped in alternation and in turn produce rivalry. This effect is however, qualitatively different from that produced by overlapped faces in that it is less readily perceived and requires voluntary effort. This is a key point that serves to distinguish perceptual alternation of objects from that which is readily evident with overlapped faces. This difference is evident when one considers the fact that neither image in the overlapped house or inverted face stimuli automatically pops-out of the display.

Additional differences between face and house encoding

Our conclusion that perceived cohesiveness is an important determinant of MR is based on the idea that faces and non-face objects are encoded using information at opposite ends of a holistic-featural continuum. However, the rivalry effect produced by faces, and the absence thereof with non-face objects, may also be related to other differences in the way these two object categories are encoded. For example, MR between tilted gratings is influenced by spatial frequency information with rivalry rates declining sharply as spatial frequency is increased (Kitterle & Thomas, 1980; Mapperson & Lovegrove, 1984). Whereas non-face images are usually recognized on the basis of high spatial frequency edge information, face recognition relies more heavily on low spatial frequency information (Biederman, 1988; Liu, Collin, Rainville & Chaudhuri, 2000). Given that the spatial frequency content of our face and house images was not controlled for, it is possible that encoding of houses on the basis of high spatial frequency information contributed to their failure to produce rivalry. However, differences in spatial frequency cannot account for the finding that the MR effect produced by faces was dependent upon orientation. Furthermore, it is unclear whether recognition on the basis of specific spatial frequencies necessarily implies that only those frequencies are selected during encoding.

Other potential differences include the amount of overlap between the two images in the face vs. house stimuli. Indeed, because the shape of faces is elliptical and that of houses is geometrical, the overall area of overlap in the face stimuli may be less than that found with overlapped houses. As such, a smaller overlap may facilitate segregation of one face from the next. Another difference is that faces are bound by continuous curves whereas houses are bound by discontinuous straight lines. In addition, faces are symmetrical in shape but most houses are not. These differences could contribute to the ease with which faces are perceived and organized as wholes. Whereas these factors may be at play when perception of overlapped faces and houses is compared, they cannot account for the previously observed difference between perception of upright and inverted overlapped faces (Boutet & Chaudhuri, 2001) where stimulus properties are equivalent.

Finally, the use of tilted images may have played a role in the rivalry effect we observed. Inspection of the overlapped stimuli suggests that images that are closer to their natural vertical orientation tend to dominate over the horizontally oriented image. This may be especially relevant for houses which are never encountered in the tilted orientation. However, the tendency for one face to stand out of the overlapped stimuli remains more pronounced than in the house stimuli even after one image is in the vertical orientation, suggesting that lack of familiarity with tilted houses was not a major determinant of the effect we observed. It would nonetheless be interesting to examine the importance of grouping principles in monocular rivalry using various procedures that do not employ tilted images, such as overlapping upright images that can be differentiated on the basis of size or color.

Concluding Remarks

The results of our study suggests that the perceptual rivalry phenomenon produced by overlapped complex images arises from a limitation in the visual system's capacity to extract more than one global structure from a visual scene at any given time. Previous research supports the importance of Gestalt grouping principles in binocular rivalry effects and ambiguous figures. It has been shown that images that are readily and unambiguously organized into a cohesive whole tend to dominate over less cohesive images (e.g., an upright vs. inverted camouflaged Dalmatian figure) under binocular rivalry conditions (Alais & Blake, 1999; Engel, 1956; Yu & Blake, 1992). There is also evidence that binocular rivalry can be driven not only by eye dominance (Blake, 1989; Tong & Engel, 2001), but also by pattern coherency (Bonneh et al., 2001; Kovacs et al., 1996). For example, competing dichoptic image fragments that share a common attribute (e.g., color) will tend to be grouped together and co-vary in perceptual dominance (Kovacs et al., 1996). Together, these findings suggest that grouping principles play an important role in binocular rivalry.

We have proposed in this paper that similar mechanisms operate during the perception of overlapped complex figures. Hence, images that are perceived as Gestalts, such as photographic positive and negative faces, will readily produce a rivalry effect whereby either face is perceived as a cohesive whole in alternation. In contrast, images that are processed in a feature-based manner, such as inverted faces and houses, do not readily produce a rivalry effect when overlapped but rather are perceived as an ambiguous mix of segregated features. This observation suggests that competition for visual awareness is not only driven by rivalry between monocular inputs but also by the perceived cohesiveness of competing images.

Preface

The hypothesis that forms the basis for the work presented in Chapter 4 is that face processing is a strong candidate for modularity and as such that faces may be used to investigate aspects of modularity that Fodor does not specifically discuss. Two questions are examined. First, whether mandatory operations bestow an attentional advantage to face processing is considered. Second, whether different modules share a common attentional pool, or have their own dedicated resources, is investigated. The results suggest that faces hold an attentional advantage during the early stages of visual processing in that they capture more attention, and are more automatically processed, than other objects. The results also indicate that faces and non-face objects compete for attention, suggesting that modular operations are subject to the same cognitive limitations as other types of processes. These findings suggest that modular systems are not fully independent because devoting resources to one module limits other modules' access to these resources. The findings also suggest that attention is a generalized resource that has extensive connections with various modular processes.

Chapter 4

Manuscript IV

Attentional competition between face and non-face objects

Boutet, Borrmann & Chaudhuri. Submitted to *Journal of Experimental Psychology: Human perception and performance.*

IV.1 Abstract

Three separate experiments were conducted to investigate whether faces and non-face objects compete for attention and if attention has a differential influence on recognition of these two stimulus types. All experiments followed the same general procedure whereby face and house images were shown during encoding. Attention was allocated to either category by way of a behavioral task. Recognition of both attended and unattended images was subsequently

measured. We conclude from all three experiments that faces and non-face objects compete for attention. Furthermore, we have found evidence that faces are more automatically encoded and grab more attention than non-face objects, but that this differential effect is evident only under very short presentation times. We conclude that faces serve as an exogenous cue to attract attention in the early stages of visual processing.

IV.2 Introduction

There is now convincing evidence that faces and non-face complex objects are analyzed by functionally and anatomically distinct mechanisms. Although there has been much research on visual processing of each stimulus class, currently there is little known about the way in which selective attention to one affects processing of the other. We address here two related aspects of this issue. First, we investigated whether or not faces and non-face objects compete for attention. Second, we examined whether or not the differential processing of faces and objects is reflected in the way attention influences their recognition.

Processing of face and non-face objects

There is now sufficient evidence that faces are processed by the human visual system in a different manner than other complex stimuli (see reviews by Tovée, 1998; Biederman & Kolacsai, 1998). Support for this notion arises from behavioral investigations that indicate perception and recognition of faces to be more greatly affected by a variety of experimental manipulations than that of other objects. Furthermore, current data point to the existence of an anatomically segregated cortical region in higher primates that is specialized for processing faces but not other complex stimuli.

A particularly robust functional difference between face and object recognition occurs through the inversion effect whereby upside-down faces are disproportionately more difficult to recognize than upside-down objects (Yin, 1969; Valentine, 1988). This effect is believed to arise from a difference in the type of information used for recognizing these two stimulus categories. There is converging evidence that faces and objects are encoded using information at either end of a featural/configural continuum (Kemp, McManus & Pigott, 1990; Rhodes, Brake & Atkinson, 1993; Sergent, 1984; Tanaka & Farah, 1993; Tanaka & Sengco, 1997; Young, Hellawell & Hay, 1987) (for a review of different definitions provided for these terms see Farah, Tanaka & Drain, 1995). In some theoretical models, face encoding is characterized as *holistic* and refers to the idea that such stimuli are processed as a whole unit or Gestalt (Ellis, 1975; Farah et al., 1995a). The difficulty to recognize inverted faces is thought to arise from an inability to access configural information such that inverted faces may instead be analyzed using a feature-based strategy similar to that applied to non-face objects. In addition to inversion, other manipulations have been shown to have a greater impact on face than object recognition. These include contrast reversal or photographic negation (e.g., Bruce & Langton, 1994; Galper 1970; Gauthier, Williams, Tarr & Tanaka, 1998; Liu & Chaudhuri, 1997), variations in lighting direction (e.g., Enns & Shore, 1997; Hill & Bruce, 1993, Johnston, Hill & Carman, 1992), and rotation in depth (e.g., Bruce, Valentine & Baddeley, 1987; Schyns & Bülthoff, 1994; Troje & Bülthoff, 1996). All of these manipulations support the idea that faces are treated differently than non-face stimuli.

Functional differences between faces and non-face objects are coupled with differences in the neuronal mechanisms that underlie their processing. Electrophysiological studies have shown that neurons in monkey area IT (inferotemporal cortex) are selective for either faces or complex object features but not to both (Perrett, Hietanen, Oram & Benson, 1992; Rolls, 1992; Wallis & Rolls, 1997; see review by Gross, 1992). Functional imaging studies in humans indicate that a specific region in the right fusiform gyrus, known as the fusiform face area (FFA), is selectively activated by faces (Kanwisher, McDermott & Chun, 1997; McCarthy, Puce, Gore & Allison, 1997; Puce, Allison, Gore & McCarthy, 1995; Tong, Nakayama, Moscovitch, Weinrib & Kanwisher, 2000). Other object categories, such as chairs and buildings, appear to activate discrete brain regions located in the vicinity of the FFA (Aguirre, Zarahn & D'Esposito, 1998; Epstein & Kanwisher, 1998; Ishai, Ungerleider, Martin, Maisog & Haxby, 1997). Finally, the notion that separate neuronal compartments are responsible for the selective processing of faces can be drawn from the clinical literature. The separate inability to recognize faces (prosopagnosia) and other equally complex object categories provides striking evidence for the existence of a specialized face-processing module (e.g., Farah, Wilson, Drain & Tanaka, 1995; Moscovitch, Winocur & Behrmann, 1997; McNeil & Warrington, 1993).

Despite the overwhelming evidence that faces and non-face objects are processed differently, the actual basis for that difference continues to be disputed. The mechanisms that are specialized for encoding faces may exist because such stimuli are endowed with a unique social and biological meaning (Farah, 1992; Farah, Wilson, Drain & Tanaka, 1998). Alternatively, faces may engage an expert subordinate-level recognition system that mediates the special processing mechanisms that have been attributed to faces. In support of the expertise hypothesis, many face-specific effects have been observed for non-face objects among individuals who had become experts in recognizing such stimuli (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Gauthier et al., 1998; Gauthier, Tarr, Anderson, Skudlarski & Gore, 1999; Gauthier, Skudlarski, Gore & Anderson, 2000). Regardless of the actual foundation for face-specific mechanisms, it is generally accepted that a functional substrate exists for the selective processing of faces.

Attention

A relatively new direction in face research concerns the role that attention plays in recognition performance and cortical activation. It has been known for some time that attention plays a crucial role in visual perception by reducing processing load through selection of a subset of the available information for greater scrutiny at the expense of irrelevant information. Attention can be allocated on the basis of physical location (see reviews by Posner & Petersen, 1990; Colby, 1991), object identity (Desimone & Duncan, 1995; Duncan, 1984; Neisser, 1967; O'Craven, Downing & Kanwisher, 1999), and intrinsic features (Corbetta, Miezin, Dobmeyer, Shulman & Petersen, 1991).

Several differences in the way selected and ignored objects are processed have been proposed. When two overlapped line-drawn objects are shown at the encoding stage (prime stimulus), subsequent presentation of test stimuli (probes) depicting the selected object are named more quickly than probes depicting new objects. In contrast, probes depicting the ignored object are named more slowly than probes depicting new objects. Greater RTs to selected primes and lower RTs to ignored primes are referred to as positive and negative priming, respectively.

This effect has been attributed to facilitation of attended object representations. Concurrently, it is believed that unattended object representations are inhibited to avoid interference with the processing of the competing attended object (Tipper, 2001). This interpretation is supported by electrophysiological and imaging studies. Neuronal responses to a preferred stimulus can be suppressed if another stimulus present within the cell's receptive field is selected (Moran & Desimone, 1985). In addition, selective attention has been shown to increase activity in those cortical areas that process the attended information and to suppress activity in areas that process unattended information (Braitman, 1984; Colby & Goldberg, 1999; Corbetta et al., 1991; Motter, 1994).

Few studies have examined whether or not objects that are encoded by distinct functional mechanisms compete for attention. Faces are ideal stimuli to test this possibility because they are believed to be encoded by distinct functional and anatomical mechanisms. Whereas many functional imaging studies have shown attentional modulation in the putative face area (Clark et al., 1997; Eimer, 2000; Haxby et al., 1994), only a few have specifically investigated competition between faces and non-face objects. O'Craven et al. (1999) have provided evidence that face and house stimuli compete for attention. However, the fact that either the house or the face was always moving in their displays raises the possibility that motion, which is a salient feature that easily captures attention, produced the face/house competition effects observed. Wojciulik, Kanwisher and Driver (1998) also observed attentional modulations in the FFA as a result of competition between faces and houses. However, because the presentation format of the images varied across testing sessions, one cannot determine whether the difference found was due to the simultaneous presence of faces and houses, or to a difference between the presentation formats of the images.

Thus, the only two studies to address this issue do not provide conclusive evidence as to whether or not attention can be separately allocated to faces and non-face objects. Given our current views on the modular basis of higher visual function, this question remains pivotal because of its implications for the role that higher cognitive factors may play across separate functional processes. An important related question concerns whether the differential processing of objects and faces itself affects the allocation of attention to each respective stimulus category.

The present study

We conducted three separate experiments to address these issues. Houses were used as a non-face category because they provide a good comparison for faces. Both types of stimuli share a set of common features that are arranged in a specific manner that renders each exemplar unique. There is also evidence that houses are processed differently from faces at the functional (e.g. Tanaka & Farah, 1993; Tanaka & Sengco, 1997) and anatomical level (Aguirre et al., 1998; Epstein & Kanwisher, 1998; Kanwisher et al., 1997; O'Craven et al., 1999).

All experiments followed the same general procedure whereby face and house images were shown during encoding. Attention was allocated to either category by way of a behavioral task. Recognition of both attended and unattended images was subsequently measured, allowing us to test for main effects of attention (attended vs. unattended), stimulus type (faces vs. houses), and for an interaction between these two variables. We speculated that attention may modulate encoding of faces vs. houses in four possible ways.

The first possibility is that faces and objects compete for attention but that faces are more automatically encoded than objects. A number of studies suggest that this might be the case. Vuilleumier (2000) and Vuilleumier and Schwartz (2001) have shown that patients with spatial neglect are less likely to miss faces than objects and scrambled faces, suggesting that faces may be processed in a very efficient and automatic fashion. Studies of "inattentional blindness" in normal observers have shown that faces can be detected despite inattention (Mack & Rock, 1998). The finding that masked faces are more easily detected than other objects also suggests an early-detection advantage for faces (Purcell & Stewart, 1988). Finally, whereas attention is essential for binding features of simple objects (Treisman, 1988; Treisman & Gelade, 1980; Treisman & Sato, 1990), there is some evidence that encoding of a face as a holistic configuration requires little attention (Boutet, Gentes-Hawn & Chaudhuri, 2000). If faces are more automatically encoded than other complex objects, then one would expect face recognition to be less influenced by attention than object recognition. Thus, automatic encoding of faces should create little interference in house encoding in our experiments. This possibility would be supported by a significant interaction between attention and stimulus type, with the difference between recognition of attended and unattended faces being less pronounced than that between recognition of attended and unattended houses.

The second possibility is that faces and objects compete for attention but that faces have an advantage over objects in capturing attention. The finding that faces are less subject to inattentional blindness (Mack & Rock, 1998) and extinction in spatial neglect (Vuilleumier, 2000; Vuilleumier & Schwartz, 2001) than other stimuli may instead suggest that faces tend to capture attention more readily than other objects. There is some evidence that the emotional content of faces can grasp attention (Bradley et al., 1997; Vuilleumier & Schwartz, 2001). Finally, preferential looking for faces in young infants is also consistent with the idea that faces hold a special status in capturing attention (Johnson, Dziurawiec, Ellis & Morton, 1991). If faces have an advantage over objects in capturing attention, then one would expect face encoding to interfere with object encoding irrespective of whether attention is devoted to the face or to the house. Finding an interaction between attention and stimulus type, with the difference between recognition of attended and unattended faces being less pronounced than the difference between recognition of attended and unattended objects, would support this possibility. Note that the idea that faces are more automatically encoded than objects is hard to dissociate from the idea that they have an advantage in capturing attention since such a finding could be interpreted either way. The same applies to the studies on extinction by Vuilleumier and colleagues (2000; 2001) and those on inattentional blindness by Mack and Rock (1998).

The third possibility is that faces and objects do compete for attention but that neither is more automatically encoded nor attracts more attention than the other. Most studies that have investigated whether faces pop out from distractors in search tasks have failed to find any advantage for faces (Kuehn & Jolicoeur, 1994; Nothdurft, 1993), suggesting that they do not hold any special status with respect to attention. If faces and objects compete for attention and if both are similarly influenced by this competition, then one would expect to find a main effect of attention without a significant interaction.

The fourth possibility is that faces and objects draw upon separate attentional resources and as such do not compete for attention. Although this might seem unlikely, the fact that attentional resources can sometimes be independently drawn across two different modalities (Mondor & Amirault, 1998) raises the possibility that visual attention might be allocated independently to different functional modules. A study by Khurana (2000, Exp. 5) provides some indication that this might be the case. Using a priming paradigm, she found that upside-down face distractors do not compete with upright target faces. Given that encoding of inverted faces closely resembles that of objects (Tanaka & Farah, 1993; Tanaka & Sengco, 1997), this study raises the possibility that the face module has its own dedicated attentional resources that are not shared with other modules. If true, then one would expect to find no main effect of attention in our study.

In the following three experiments, we have explored this question by use of overlapped transparent stimuli, spatially separated stimuli, and a priming paradigm. We conclude from all three experiments that faces and non-face objects compete for attention. Furthermore, we have found evidence that faces are more automatically encoded and attract more attention than non-face objects, but that this differential effect is evident only under very short presentation times.

IV.3 Experiment 1A

Competition between upright overlapped faces and houses

In Experiment 1, we examined whether faces and objects compete for attention using overlapped stimuli created by superimposing a face on a house with 50% transparency. Neisser and Becklen (1975) have shown that overlapped transparent images are a valid tool for the study of selective attention. The use of such images involves the deployment of an object-based type of attention since the images are presented at the same location. A similar object-based paradigm has been successfully used to illustrate attentional modulation in face- and object-processing areas (O'Craven et al., 1998). With overlapped displays, visual inspection of both images is possible and attentional focus cannot be confounded with foveation. This paradigm also controls for eccentricity confounds found in some spatial attention studies where images are shown at either side of fixation.

During presentation of each overlapped stimulus, participants were instructed to use a mouse to manually delineate the face (face attended condition), the house (house attended condition), or both (divided attention). Recognition performance was measured for both faces and houses at testing. If faces and houses do indeed compete for attention, then recognition of both faces and houses should be greater when testing for the attended stimulus category than for the unattended stimulus category. If faces and houses do not compete for attention, then recognition should be comparable in the attended and unattended conditions.

IV.3.1 Method

Participants

Thirty-nine (7 male) undergraduate students from McGill University took part in this experiment. All participants were naïve with regard to the goals of the experiment. Their ages ranged from 18 to 25. All participants had normal or corrected-to-normal vision.

Stimulus Materials

Eighty-eight digitized photographs of male faces were obtained from a face database at the University of Essex (http://cswww.essex.ac.uk/allfaces). The original full-color face images were converted to a 256 gray-level format. The face images were 180 x 200 pixels, subtending 6.5 x 7 degrees of arc at a viewing distance of 57 cm. Eighty-eight digitized photographs of houses were obtained from various real estate websites. The house images were scaled and cropped to fit into at 180 x 200 pixels window and converted to a 256 gray-level format. All images were unknown to the participants. The stimuli were equalized to an average luminance of 49.2 cd/m^2. Sixty-four faces and 64 houses were randomly selected from a pool of 88 to construct the encoding stimuli. The selected face and house images were paired and overlapped with 50% transparency using Adobe Photoshop 3.0.5 software. The remaining faces and houses were used as testing distractors.

Participants were individually tested using a Macintosh G3/266 computer. The stimuli were presented on a 21" Sony color monitor. The screen was calibrated for luminance using an Optikon Universal Photometer. Images were surrounded by a neutral gray background of 44.35 cd/m^2.

Procedure

This study employed a sequential matching paradigm. A trial consisted of an encoding stage followed by a testing stage (Figure 1). Two overlapped stimuli were sequentially shown during the encoding stage—one was a target and the other a dummy. Prior to stimulus presentation, participants were instructed on the screen whether to select the face (attended face), the house (attended house), or both (divided attention). When the stimulus was presented, participants had to delineate the appropriate structure with the aid of a mouse. Trials where the house was delineated in the target stimulus and the face in the dummy stimulus are referred to as 'attended house' trials. For the attended house trials, recognition accuracy was evaluated for the attended target house as well as for the

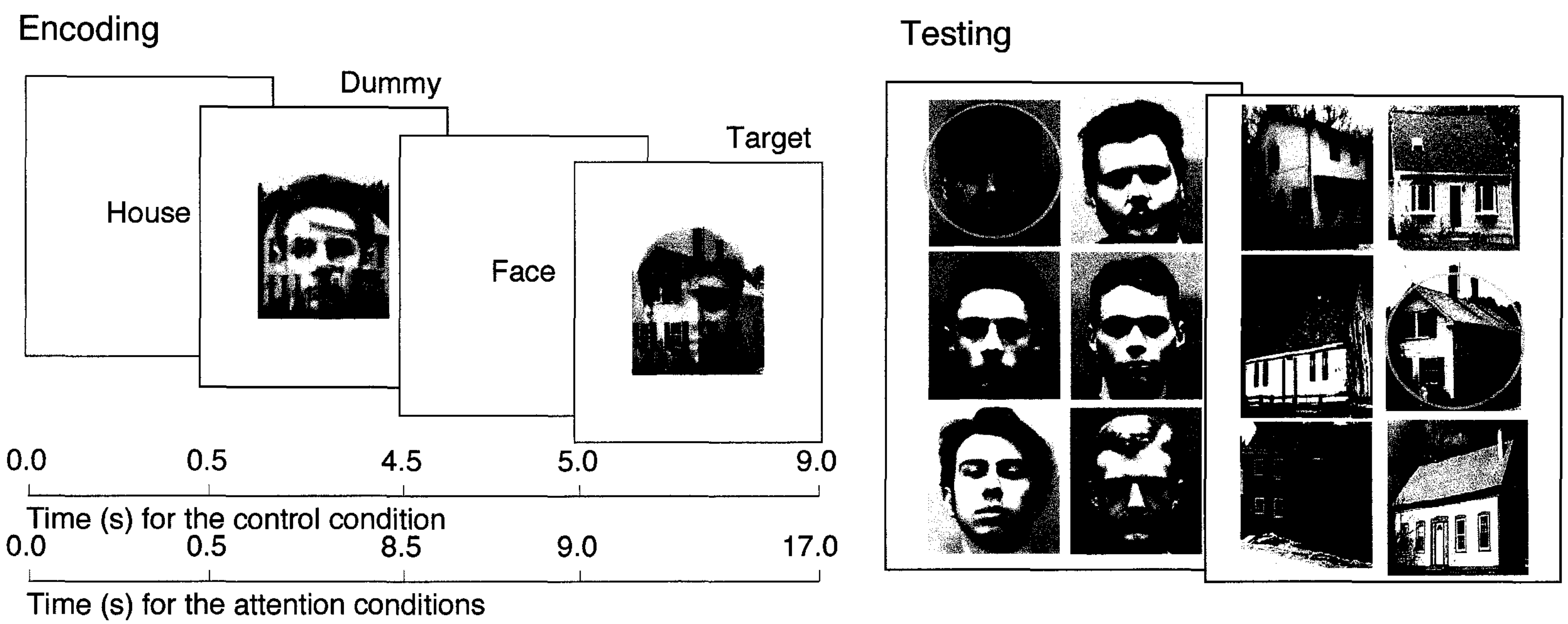

Figure 1. A schematic illustration of the sequential matching procedure that was used in Experiment 1. In the trial illustrated here, recognition performance was assessed for an attended face and an unattended house. Correct responses are circled.

overlapping unattended face. Trials where a face was delineated in the target stimulus and the house in the dummy stimulus are referred to as 'attended face' trials. For the attended face trials, recognition accuracy was evaluated for the attended target faces as well as for the overlapping unattended house. In divided attention trials, participants had to delineate the conjunction of the face and the house for both the target and the dummy stimuli. Recognition was evaluated for both the face and the house presented in the target stimulus. Hence recognition of the face and house images shown in the dummy stimulus was never evaluated.

We evaluated the accuracy of object delineation by each participant in the following manner. We recorded 100 pixel values corresponding to the area on the screen being delineated during the 8s duration of each frame. A delineation performance score was calculated by comparing each one of these 100 pixel values to eight pixel values corresponding to the outline of the face, the house, or the conjunction of both. A score of 1 was given for any pixel value that was within ± 20 pixels of any one of these eight pixel values. The data obtained from participants who scored more than two standard-deviations apart from that of all other participants was to be discarded. Participants were instructed to devote all of their attention to what they were delineating. After the experiment, participants were asked if they felt confident that they did so; data from participants who doubted their ability to focus on the image to be delineated were discarded.

Two arrays were shown sequentially during the testing phase of each trial—one made up of 6 faces and the other of 6 houses. One face or house was the same as that shown in the target encoding stimulus. The other faces or houses were randomly chosen among the testing distractors. A given distractor could be shown a maximum of three times at testing. Participants had to select which face and house seem familiar. Recognition accuracy (%) and median reaction times (RT) for correct responses were recorded. The presentation order of the face and house arrays was counterbalanced.

We used this modified sequential matching paradigm to ensure that participants fully attended to the relevant stimulus rather than developing a strategy of divided attention. This would have been likely if a single encoding trial was presented because participants would quickly realize that recognition performance for both the face and house were being assessed. The dummy stimulus therefore served to minimize guessing which of the two stimuli shown at encoding contained the image to be used for later testing. The presentation order of the target and dummy stimuli was counterbalanced to additionally ensure that participants did not guess which served as the actual target.

The procedure for the control condition was identical except that no attentional instructions were shown and participants passively viewed the stimuli. The stimuli remained on the screen for 4 s in the control condition trials. Participants were first tested on a session of eight trials for the control condition, and then tested on a random presentation of eight trials for each of the remaining conditions. The 64 overlapped stimuli were randomly assigned to either target or dummy and to one of the four experimental conditions.

Preliminary Experiment

We conducted a preliminary experiment to measure baseline recognition of the 64 houses and faces that composed the overlapped stimuli. Face and house recognition was measured separately (counterbalanced order across participants) using an "old/new" recognition paradigm in ten new participants (3 males), whose ages ranged from 22 to 34. In the encoding stage, 32 faces or 32 houses were randomly selected from the pool of 64 and shown sequentially for 4 s. In the testing stage, the 32 encoding stimuli were shown sequentially in a random order amongst the 32 remaining images as distractors. Whether or not each image was shown during encoding had to be determined. The average face and house recognition accuracies obtained were 80.49% (*SD* 7.67) and 70.31% (*SD* 11.81), respectively. Thus, the baseline difference between face and house recognition was approximately 10%. Median RTs for correct responses were 0.95 s (*SD* 0.17) and 1.241 s (*SD* 0.29), respectively.

IV.3.2 Results

Data from two participants were discarded because they reported having difficulties directing their attention on the image they were delineating. No data were discarded because of poor performance on the delineation task.

Accuracy

We found that faces were better recognized than houses overall (Figure 2A). Attention had a similar influence on these two types of stimuli because both faces and houses were best recognized when selected and least recognized when ignored. Finally, face and house recognition performance was intermediate in the divided attention condition.

A 3 x 2 repeated-measures ANOVA with Attention (attended, unattended, and divided) and Stimulus Type (face and house) as variables was performed on the average recognition accuracy. The main effects of Attention [$F(2, 72) = 19.73$, $p < .01$] and Stimulus Type [$F(1, 36) = 22.48$, $p < .01$] were significant. The Attention X Stimulus Type interaction was not significant [$F(2, 72) = 0.56$, $p = .57$]. Because the presentation time of the learning stimuli was much shorter for the control condition, the data obtained in that condition were analyzed separately. A *t* test [$t(36) = 18.15$, $p < .01$] revealed that passive viewing of the overlapped stimuli yielded better recognition of the faces (*M* 66 *SD* 22) than the houses (*M* 49 *SD* 21).

RT

The reaction time (RT) responses were consistent with the recognition accuracy data. The same statistical analyses were performed on the median RTs for correct responses. Data from those participants who obtained 0% accuracy in one of the attention condition tested were discarded. The main effects of Attention [$F(2, 62) = 7.85$, $p < .01$] and Stimulus Type [$F(1, 31) = 14.53$, $p < .01$] were significant. The interaction approached significance [$F(2, 62) = 3.03$, $p = .06$]. As can be seen in Figure 2B, this effect can be attributed to the attended and divided conditions producing a greater

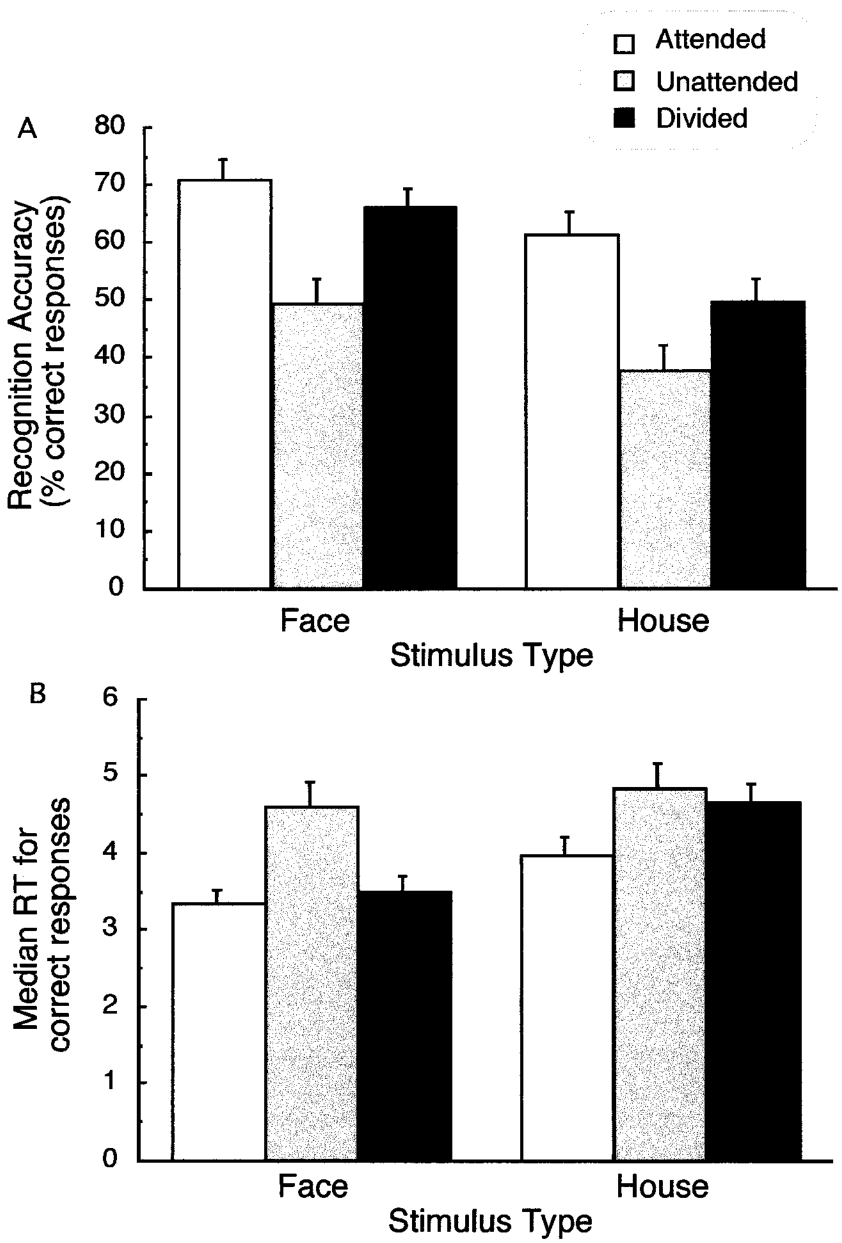

Figure 2. **A** Average face and house recognition accuracy (37 Ss), and **B** Average median RT for correct responses (32 Ss) obtained in Experiment 1A for the attended, unattended, and divided attention conditions. Error bars represent ± 1 S.E.

performance for the houses than for the faces, and the reverse holding for the unattended condition. This finding has little bearing on the hypotheses tested because faces and houses display a similar trend across attention conditions.

The median RT for correct responses obtained in the control condition was analyzed separately. Data from those participants who obtained 0% accuracy in one of the control condition tested were discarded. A *t* test [t (36) = 0.09, p = .77] revealed that faces (3.71 s) and houses (3.78 s) yielded the same response latencies under passive viewing condition. The discrepancy between the accuracy and RT data for the control condition is probably due to the relatively difficult nature of the task. Indeed, RT is a more sensitive measure when the task is relatively easy and the level of accuracy is high. In difficult tasks, however, accuracy data are more relevant because differences in RT are more easily obscured by variability.

IV.3.3 Discussion

Our results indicate that faces and objects compete for attention. Recognition performance was found to be best for attended images and worst for unattended ones. Recognition in the divided attention condition occurred somewhere between that for attended and that for unattended images. The greater recognition performance in the attended condition may likely be due to an enhancement of the neuronal networks probed by the attended image, as shown previously in electrophysiological and imaging studies (Braitman, 1984; Colby & Goldberg, 1999; Corbetta et al., 1991; Moran & Desimone, 1985; Motter, 1994). Similarly, suppression of neural networks probed by the unattended images may account for the poor recognition performance in that condition. The implication of putative facilitating and inhibiting mechanisms is further discussed in Experiment 3.

The results of Experiment 1A suggest that allocating attention to a given object category interferes with the processing of another object category. Assuming that faces are processed by a different functional and anatomical module than houses, these findings imply that attention is not module-specific but rather exists as a generalized resource that is allocated to multiple aspects of visual processing. The absence of a significant Attention X Stimulus Type interaction further suggests that faces and non-face objects are similarly influenced by attention despite differences in the way they are encoded. Indeed, if faces have an advantage over objects in attracting attention, then one would expect face recognition in the divided attention condition to be equal to that in the full attention condition. Our results did not show this. Alternatively, if faces are more automatically encoded than houses, thus requiring less attention, then we would have expected face recognition to be less affected when attention was devoted to the houses than house recognition to have been affected when attention was devoted to the faces. It is clear from Figure 2 that this was not the case. Instead, the same pattern of performance was found for both stimulus categories.

In the preliminary experiment, where the images were shown in a normal non-overlapped fashion, faces yielded 80% recognition accuracy compared to houses at 70%. In Experiment 1A, passive viewing of the overlapped stimuli for the same time yielded a recognition accuracy of 66% for the faces and 50% for the houses. Although the difference between face and house recognition is comparable (~10%), it is clear that overlapping the stimuli affected visual encoding as reflected by the poorer recognition performance. This difference raises the possibility that the technique of overlapping transparent stimuli degrades image properties that are inherent to faces and as such disrupts perception of a face as a face. We examined this possibility in Experiment 1B by testing for an inversion effect with our overlapped displays.

IV.4 Experiment 1B

Competition between inverted overlapped faces and houses

Measuring the effect of inversion on face vs. object recognition is a standard way to test for the "faceness" of a stimulus, given that the inversion effect is among the most robust of behavioral differences between faces and non-face objects (Gauthier & Tarr, 1997; Valentine, 1988). Consistent with the inversion effects that have been previously reported, we predicted that face recognition would be more greatly affected by inversion than house recognition. The same procedures as that used in Experiment 1A were employed in Experiment 1B, except that all images were inverted.

IV.4.1 Method

Participants

Thirty-seven students (6 male) from McGill University took part in this experiment. All participants were naïve with regard to the goals of the experiment. Their ages ranged from 18 to 28. All participants had normal or corrected-to-normal vision.

Stimulus Materials and Procedure

Same as Experiment 1A, except that all the images were inverted.

IV.4.2 Results

Data from one participant were discarded because of difficulties in directing attention on the image to delineate. No data were discarded because of poor performance on the delineation task.

Accuracy

The results of Experiment 1B are consistent with those of Experiment 1A (Figure 3A). Attention had a similar influence on these two types of stimuli since both faces and houses were best recognized when selected and least recognized when ignored. However, the superior face recognition performance obtained in Experiment 1A no longer holds here. Rather, recognition performance for the faces and houses is similar.

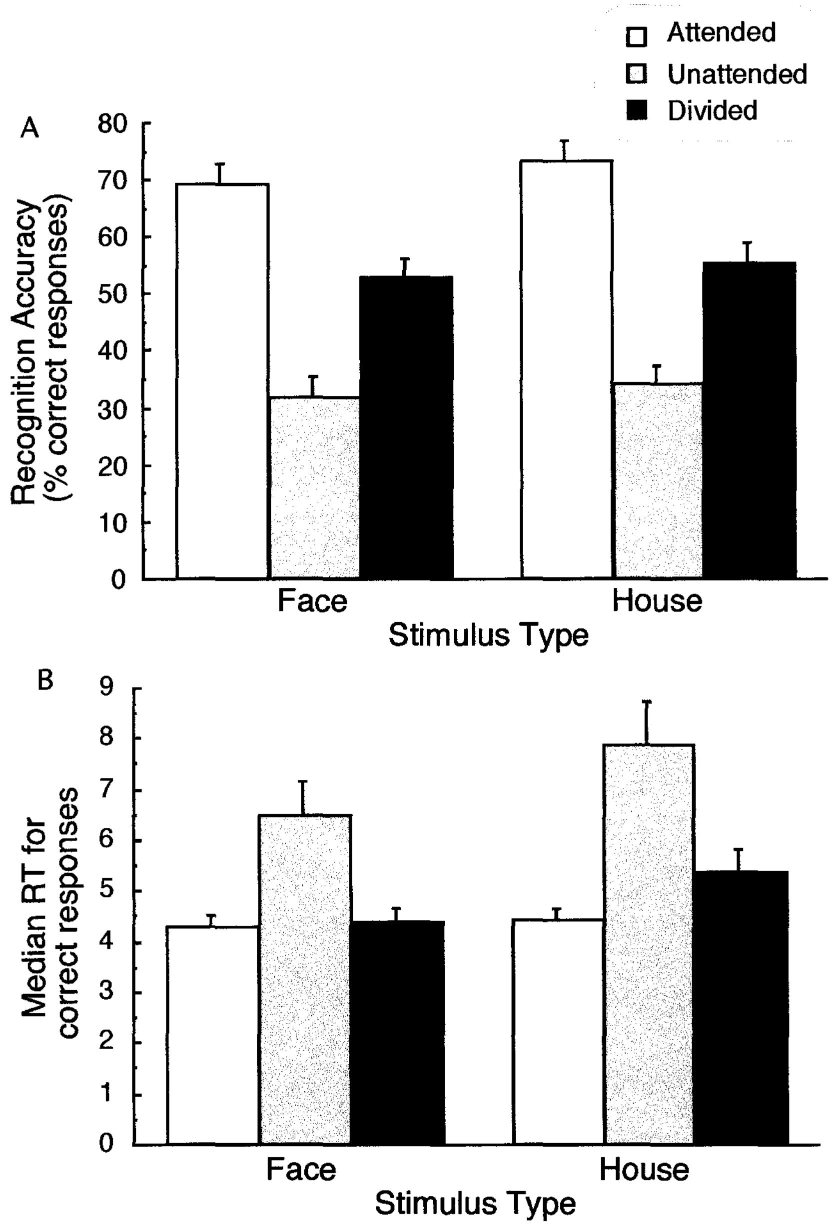

Figure 3. **A** Average face and house recognition accuracy (37 Ss), and **B** Average median RT for correct responses (32 Ss) obtained in Experiment 1B for the attended, unattended, and divided attention conditions. Error bars represent ± 1 S.E.

A 3 x 2 repeated-measures ANOVA with Attention (attended, unattended, and divided) and Stimulus Type (face and house) as variables was performed on the average recognition accuracy. The main effect of Attention was significant [$F(2, 72) = 67.62$, $p < .01$]. The main effect of Stimulus Type [$F(1, 36) = 2.03$, $p = .16$] and the Attention X Stimulus Type interaction were not significant [$F(2, 72) = 0.06$, $p = .94$]. A *t* test [$t(36) = 1.17$, $p = .29$] revealed that passive viewing of the overlapped stimuli yielded comparable recognition of faces (*M* 46 *SD* 20) and houses (*M* 51 *SD* 19). The fact that the effects of stimulus type disappeared after inversion indirectly supports an inversion effect.

A direct test of an inversion effect was obtained by comparing the results of Experiments 1A and 1B using a mixed-design ANOVA with Orientation (upright and inverted), Attention, and Stimulus Type as variables. The Orientation X Stimulus Type interaction was significant [$F(1, 36) = 21.15$, $p < .01$], supporting the notion that inversion disproportionately affected face recognition as compared to house recognition. The overall difference between recognition of upright and inverted faces was 11%.

RT

The same statistical analyses were performed on the median RTs for correct responses (see Figure 3B). Data from those participants who obtained 0% accuracy in one of the conditions tested were discarded. Data from another participant were discarded because of a computer error. The main effects of Attention [$F(2, 62) = 17.62$, $p < .01$] and Stimulus Type [$F(1, 31) = 5.68$, $p = .02$] were significant. The interaction was not significant [$F(2, 62) = 1.75$, $p = .18$]. Comparison of Experiments 1A and 1B indicates that the Orientation X Stimulus Type interaction was not significant for the RT data [$F(1, 62) = 0.13$, $p = .72$]. A *t* test [$t(34) = 0.279$, $p = .60$] revealed that passive viewing of the overlapped stimuli yielded comparable RT for faces (*M* 4.52 *SD* 1.90 s) and houses (*M* 4.69 *SD* 1.76 s).

IV.4.3 Discussion

Our results indicate that face recognition was more greatly affected by inversion than house recognition. This finding suggests that the faces in the overlapped upright stimuli retained their facial quality and were processed as faces. The absence of a significant inversion effect (i.e., Orientation X Stimulus Type interaction) with the RT data does not undermine this conclusion given that this effect usually holds for recognition accuracy measures (Gauthier & Tarr, 1997; Valentine, 1988). As in the previous experiment, the difficult nature of the recognition task may account for discrepancies between the accuracy and RT measures. Given that the overlapped stimuli yielded an inversion effect, we assume that Experiment 1A successfully tested for encoding and recognition of face and non-face images that probe distinct functional and anatomical mechanisms. We therefore retain our prior conclusion that faces and non-face objects compete for attention.

Inverting the stimuli served another purpose in that it allowed us to further examine whether differences in the type of information used to analyze faces and non-face objects are reflected in the way they are influenced by attention. Indeed, there is converging evidence that inversion disrupts holistic face encoding such that inverted faces may instead be processed in a feature-based manner similar to that applicable to non-face objects (Rhodes et al., 1993; Sergent, 1984; Tanaka & Farah, 1993; Tanaka & Sengco, 1997;Young et al., 1987). Assuming that faces are encoded holistically, and that inversion disrupts this process, one can compare the results of Experiment 1A and 1B as a way to examine the allocation of attention when holistic and feature-based encoding processes compete vs. when feature-based processes alone compete.

Consistent with Experiment 1A, the results of Experiment 1B indicate that inverted faces and houses are similarly influenced by attention as evidenced by the absence of a significant Orientation X Attention effect. In addition, attentional modulations were almost identical across the two experiments with selected images being best recognized, ignored ones being the most poorly recognized, and recognition in the divided attention condition being intermediate. Together, these results suggest that the deployment of attentional resources is limited, irrespective of whether the competing images are processed using the same or different functional and anatomical mechanisms. Furthermore, we did not find evidence on the basis of these experiments that face encoding makes a greater or lesser demand on attentional resources than non-face objects.

IV.5 Experiment 2

Competition between spatially separated faces and houses

It may be argued that the use of overlapped images during encoding recruits low-level segregation mechanisms and that these may have been responsible for the attentional effects we observed in Experiments 1A and 1B. The use of overlapped displays has advantages over other procedures in that it allows visual presentation and processing of both attended and unattended images at the same retinal locus. However, overlapped stimuli require segregation of figure from ground and as such may involve confounding low-level processes. Thus, the effects we observed in Experiment 1 may be attributable to image segmentation processes rather than attentional selection. To further complicate matters, it is unclear whether or not such segregation mechanisms would precede or follow attentional selection (Neisser, 1967). The fact that recognition of unattended images was above chance leads us to believe that attentional selection and not low-level segregation mechanisms was involved in Experiment 1. Nevertheless, we wished to verify whether or not our findings could be replicated with non-overlapped displays.

In Experiment 2, we employed a procedure similar to that in Experiment 1 but the stimuli presented during encoding were spatially separated. We also wanted to avoid any confounds related to the use of a location-based manipulation of attention, such as foveation of attended images. In this experiment, we presented pairs of images (face and house) in a random manner at pre-defined locations. Participants were instructed to attend to the face, house, or both. The task was to note three sequential presentations of either image token at the same location. Fixation was maintained at the center of the screen and verified through eye movement recording. Furthermore, the likelihood of saccades was reduced by employing display durations limited to 200 ms. This procedure allowed us to present attended and unattended images at separate locations and control for confounds of eccentricity and reorienting eye movements.

IV.5.1 Method

Participants

Forty-seven (6 male) undergraduate students from McGill University took part in this experiment. All participants were naïve with regard to the goals of the experiment. Their ages ranged from 19 to 28. All participants had normal or corrected-to-normal vision.

Stimulus Materials

The same 88 face and 88 house images as those used in Experiment 1 were employed. An additional 32 faces and 32 houses were obtained from various internet pages and modified to match those in the original set. The stimuli were cropped to eliminate their background and equalized to an average luminance of 47.7 cd/m^2.

The same apparatus as that used in Experiment 1 was employed. Eye position was recorded for all observers using the Ober2 infrared reflection system (Permobil Meditech, Natick, MA). Eye position was sampled at 100 Hz. Head movements were reduced by use of a chin rest.

Procedure

A similar sequential matching paradigm as that used in Experiment 1 was employed. However, the stimuli were presented in a different manner during encoding (Figure 4). As in Experiment 1, a single trial was composed of one target sequence and one dummy sequence, followed by two testing arrays. Each encoding sequence consisted of the presentation of 30 frames in quick succession. Each frame in a given sequence contained the same face and house image presented at one of four randomly chosen locations (left, right, top, bottom). A fixation point appeared on the screen 2 s before presentation of the first frame. Each frame was shown for 200 ms in order to reduce saccadic eye movements. A 200 ms blank interval was presented between frames. The fixation point remained at the center of the screen for the entire sequence. Participants were instructed to fixate on that point. The experimenter used the eye movement trace during testing to provide verbal feedback to the participants and encourage them to fixate on the center when they failed to do so. Onset of the frames was accompanied by a tone. Prior to stimulus presentation, participants were instructed on the screen to count the number of times the face, the house, or both were presented at the same location in three successive frames (defined as a *triplet*). Each learning trial was correspondingly classified as *attended face, attended house*, or *divided attention*. The number of triplets was randomly chosen from two to five in any given sequence of 30 frames. There were no instances where two successive triplets were presented back to back. With the exception of the triplets, the locations of the face and house were randomly chosen across frames.

At testing, only recognition of the face and house shown in the target sequence was measured. The face and house testing arrays contained the target along with two distractors, each being randomly presented at the left, center, or right positions on the screen. Participants were instructed to select which of the three images had been shown during encoding using the time needed. Distractors were chosen from those images that were not presented during encoding. Each distractor was presented only once. Twenty-four trials were tested in total, eight per attentional condition. Each experimental session was preceded by two practice trials. No passive viewing control condition was tested in Experiment 2. Recognition accuracy (%) was recorded.

Figure 4. A schematic illustration of the sequential matching procedure that was used in Experiment 2. In the trial illustrated here, recognition performance was assessed for an attended house and an unattended face. Correct responses are circled.

IV.5.2 Results and Discussion

Data from four participants were discarded because they showed unusually frequent head movements during testing. Data from two participants were discarded because their performance on the triplet-counting task was more than two standard-deviations below the mean. Average performance on the counting task for the remaining participants was 53% (SD 22.60) for the attended face, 52% (SD 25.00) for the attended house, and 37% (SD 20.00) for the divided attention condition. Poorer performance on the divided attention condition task can be attributed to the fact that the face and the house could be repeated at the same location for two simultaneous frames as well as three simultaneous frames. Participants reported having difficulties counting triple repetitions only and ignoring double repetitions.

Accuracy

A 3 x 2 repeated-measures ANOVA with Attention (attended, unattended, and divided) and Stimulus Type (face and house) as variables was performed on recognition accuracy. The main effects of Attention [$F(2, 80) = 25.90$, $p < .01$] and Stimulus Type [$F(1, 40) = 49.21$, $p < .01$] were significant. The Attention X Stimulus Type interaction was not significant [$F(2, 80) = 2.07$, $p = .13$]. As can be seen in Figure 5, faces were better recognized than houses overall. Attention had a similar influence on these two types of stimuli since both were best recognized when selected and least recognized when ignored. Finally, whereas face recognition was higher in the divided attention than in the unattended condition, house recognition was equivalent in these two conditions. However, this difference was too small to produce a statistically significant interaction.

The results of Experiment 2 are consistent with those of Experiment 1A where faces and houses were found to compete for attention. Again, we take these findings as evidence that attention is not module-specific but instead exists as a generalized resource that is allocated to multiple aspects of visual processing. The non-significant Attention X Stimulus Type interaction found in Experiment 1A

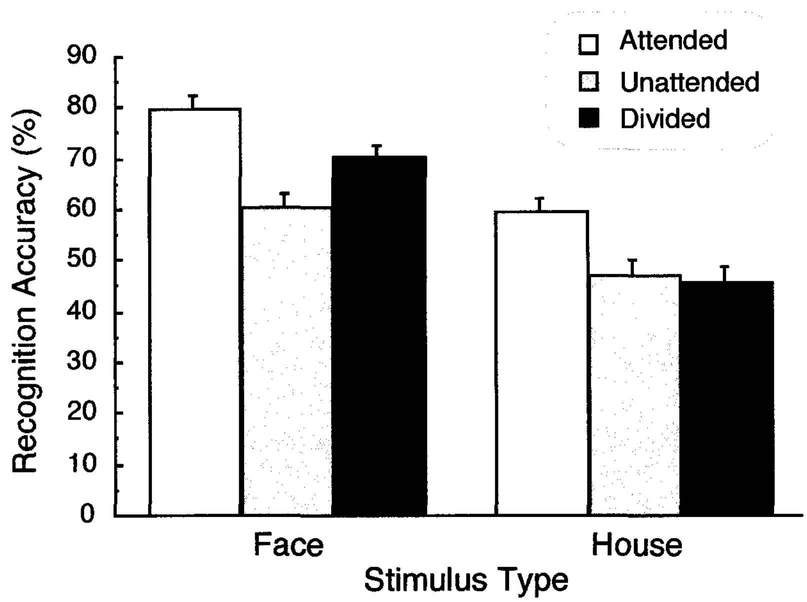

Figure 5. Average face and house recognition accuracy (41 Ss) obtained in Experiment 2 for the attended, unattended, and divided attention conditions. Error bars represent ± 1 S.E.

was also replicated. This finding suggests that faces are neither more automatically encoded than houses nor do they hold an advantage in grasping attention. Rather, it appears that faces and non-face objects are similarly influenced by attention.

Eye position monitoring

Eye movement data from 14 randomly chosen participants were analyzed with the aid of a custom computer program that used a velocity criterion to detect saccades (R. McPeek, personal communication, May 2001). The average number of saccades recorded across encoding sequences was 2.65 (*SD* 2.28). Given that 30 frames were shown in total, this analysis suggests that either one of the two images shown during encoding could have been foveated in 9% of the presented frames on average. We therefore conclude that participants successfully followed the instructions to fixate the center of the screen and that foveation of the attended information cannot account for the attentional effects we have obtained.

IV.6 Experiment 3

Priming of competing faces and houses

Our results thus far indicate that attended images are more accurately recognized than unattended ones. These results are consistent with prior physiological and imaging studies on selective attention (e.g., Corbetta et al., 1991; Moran & Desimone, 1985). These studies have shown that attending to a given image enhances its neural processing. Furthermore, the processing of unattended images is inhibited, likely to avoid interference with encoding of the relevant information.

Numerous studies have shown that unattended images are negatively primed (e.g., Tipper, 1985; Tipper and Driver, 1988; Treisman & DeSchepper, 1996), an effect characterized by slower response times and higher error scores upon presentation of a previously unattended stimulus as compared to a new stimulus. This effect has recently been shown to extend to unfamiliar face stimuli (Khurana, 2000). In contrast, positive priming yields faster response times and fewer errors upon presentation of a previously attended stimulus, including faces (e.g., Bruce, Burton, Carson, Hanna, & Mason, 1994). Whereas the mechanisms responsible for negative priming have long been disputed (see reviews by Fox, 1995; May, Kane, & Hasher, 1995), there is now compelling evidence for the involvement of inhibitory mechanisms (Tipper, 2001). As such, priming paradigms are useful in studying attentional selection processes because they permit a more direct assessment of the role of activating vs. inhibiting mechanisms. We explored this issue in Experiment 3.

The priming paradigm we adopted consisted of two tasks—a prime task and a probe task. During the prime task, face and house stimuli were presented simultaneously and participants directed their attention to one or the other stimulus category. In the probe display, subjects saw primed or unprimed images of the attended or unattended category, allowing for measures of repetition priming. In order to avoid differences in retinal eccentricity between attended and unattended stimuli, all three images in the prime task were displayed at equal distance from fixation and randomized between 4 possible locations. During trials, subjects fixated the center of the screen and attention was directed to the target images without foveation. Deliberate saccades toward the images were reduced by presenting the prime display for only 200 ms.

We reasoned that if faces and houses compete for attention, recognition of previously attended stimuli should be better than baseline (performance on 'new' probe trials), and better still than recognition of previously unattended stimuli. Positive priming should be found for attended stimuli of both stimulus categories whereas negative priming should be found for unattended face and house stimuli. If, however, faces and houses do not compete for attention, then both attended and unattended stimuli should be positively primed.

IV.6.1 Method

Participants

Thirty-seven (11 male) undergraduate students from McGill University took part in this experiment. Participants' age ranged from 18 to 36 years. All participants were naïve with regard to the goal of the study. All participants had normal or corrected-to-normal vision.

Stimulus Materials

Stimuli consisted of 288 photographs of houses and male faces (144 images per category). The same 120 face and 120 house images as those used in Experiment 2 were employed. An additional 24 faces and 24 houses were obtained from various internet pages and modified to match those in the original set. The stimuli were cropped to eliminate their background and equalized to an average luminance of 37.15 cd/m^2.

The same apparatus as in Experiments 1 and 2 was used.

Procedure

Each trial consisted of two parts—a prime task followed by a probe task (Figure 6). The prime task started with presentation of a 525 ms fixation point. Black rectangles were then shown for 75 ms at two of four possible locations of equal eccentricity from fixation. The rectangles served as pre-cues and participants were instructed to direct their attention to these two locations. This was followed by three images, two of which were presented in the two cued locations. These two target images were either houses only or faces only. The two face or house stimuli could be either identical or different. Participants were required to perform a same-different judgment on the cued targets. The third image in the prime display represented the uncued distractor. This was always drawn from a different stimulus category than the two targets (house-distractor for face-targets; face-distractor for house-targets). The two prime conditions were

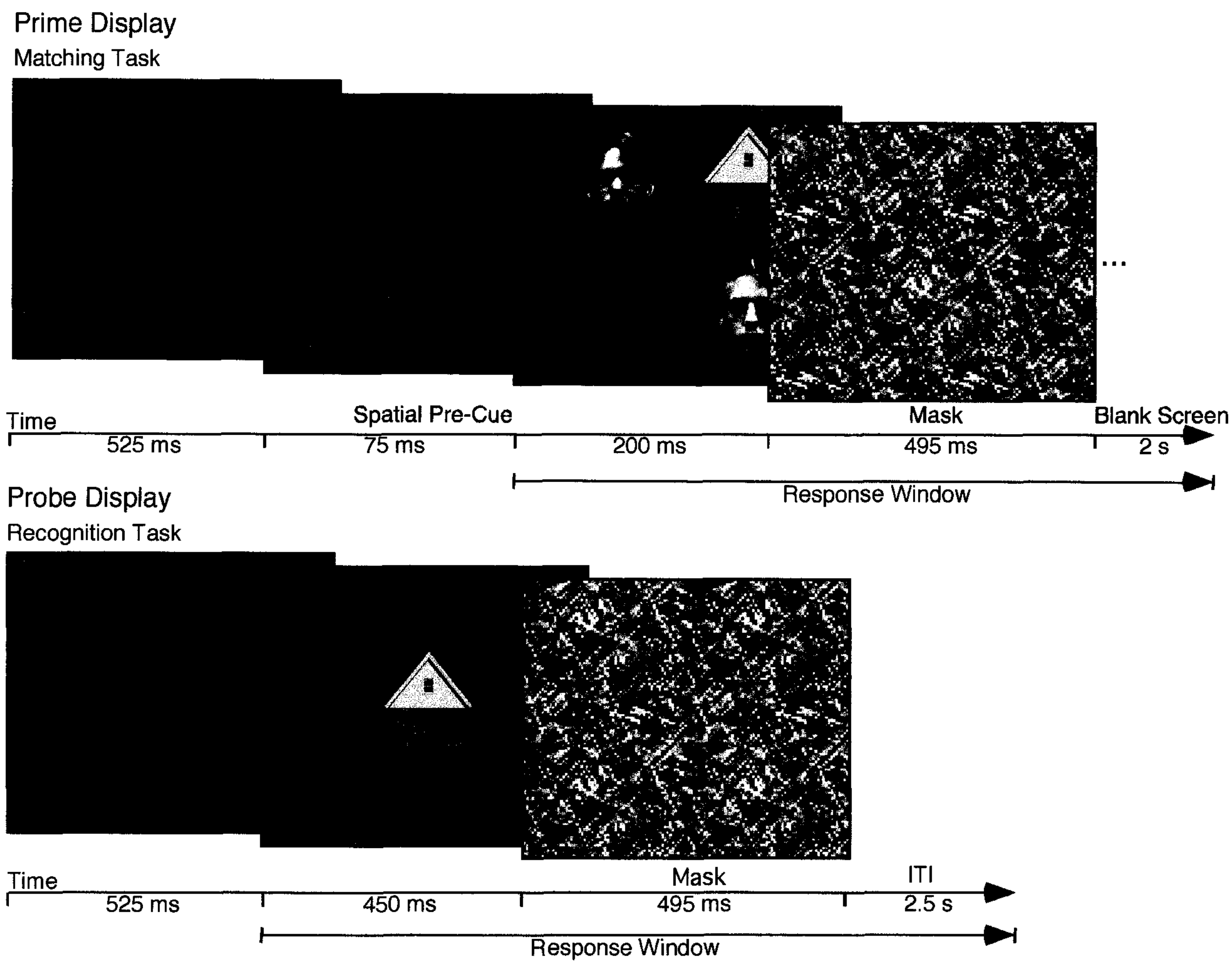

Figure 6. A schematic illustration of the priming paradigm that was used in Experiment 3. In this example, recognition performance was assessed for a previously unattended old house. Correct responses on prime and probe tasks in this example are 'same' and 'old', respectively.

house-attended and face-attended. The prime display lasted for 200 ms in order to reduce the likelihood of foveating the targets. This was followed by a 495 ms grayscale mask that filled the entire screen. The luminance of each pixel in the mask was randomly chosen from between 11.55 and 54 cd/m^2. Feedback on error was given by way of a 500ms tone.

Probe presentation was preceded by a 2 s blank interval and a 525 ms fixation point. The probe display consisted of a single image located in the center of the screen. The probe image could be one of eight possibilities—the attended stimulus category (face or house) or unattended distractor category (face or house), each of which could have been presented during the previous prime display ('old') or had never been presented before ('new'). Participants performed an old/new recognition judgment on the probe stimulus. No error feedback was given. Probe presentation lasted 450 ms, followed again by a 495 ms mask. Each participant was tested on four blocks of 24 trials each (96 trials total), preceded by 6 practice trials that were not included in the analysis. The inter-trial interval was 2.5 s.

Response times and accuracy were measured for performance both on prime and on probe task. Probe trials that followed an incorrect response on prime presentation were excluded from the analysis, because it was not certain whether subjects were paying attention and thus received priming on these trials. The magnitude of response priming was determined by subtracting each participant's response times on 'new' trials from the response times on 'old' trials.

IV.6.2 Results

Prime task

Repeated-measures ANOVA were conducted for mean correct response times (RT) and mean error rates on variable stimulus type (face vs. house). RT for the face and house matching tasks were 1.14 and 1.13 s, respectively. Both tasks yielded 90% correct responses. There was no main effect of stimulus type [$F(1, 36) = .29$, $p = .60$ for RT; $F(1, 36) = .04$, $p = .84$ for correct responses] suggesting that task difficulty was equivalent for face and house trials.

Probe task

A 2 x 2 repeated measures ANOVA with Attention (attended, unattended) and Stimulus Type (face, house) as independent variables was performed on response time priming and recognition accuracy performance, separately (Figure 7). For recognition accuracy, significant main effects of Attention [$F(1, 36) = 26.06$, $p < .01$] and Stimulus Type [$F(1, 36) = 26.06$, $p < .01$] were found. The Attention x Stimulus Type interaction was also significant [$F(1, 36) = 4.15$, $p = .05$]. This interaction is attributable to a significant difference between recognition of attended and unattended houses that was not present when comparing recognition of attended and unattended faces.

For RT priming, the main effect of Attention [$F(1, 36) = 24.67$, $p < .01$] was also significant. The main effect of Stimulus Type was not significant [$F(1, 36) = 2.93$, $p = .96$]. There was no significant Attention x Stimulus Type interaction [$F(1, 36) = 1.28$, $p = .27$]. This non-significant effect is attributable to the high variability produced by comparing positive and negative RT priming measures, but inspection of the graphs indicates the presence of an interaction effect. Matched–pair *t* tests were conducted to determine conditions in which significant priming had occurred relative to the baseline 'new' condition. These analyses revealed significant priming in the face-attended [$t(36) = -2.70$, $p < .05$], face-unattended [$t(36) = 3.88$, $p < .01$], and house-attended conditions [$t(36) = -2.88$, $p < .01$] but not the house-unattended condition [$t(36) = -.09$, $p = .93$]. Priming for previously attended faces was larger than 185 ms and priming for previously attended houses exceeded 220 ms. The significant priming effect for previously unattended faces reflects a 175 ms slowing of response times relative to new unattended faces, i.e., negative priming. As with response accuracy, no significant negative priming was found for RT to previously unattended houses.

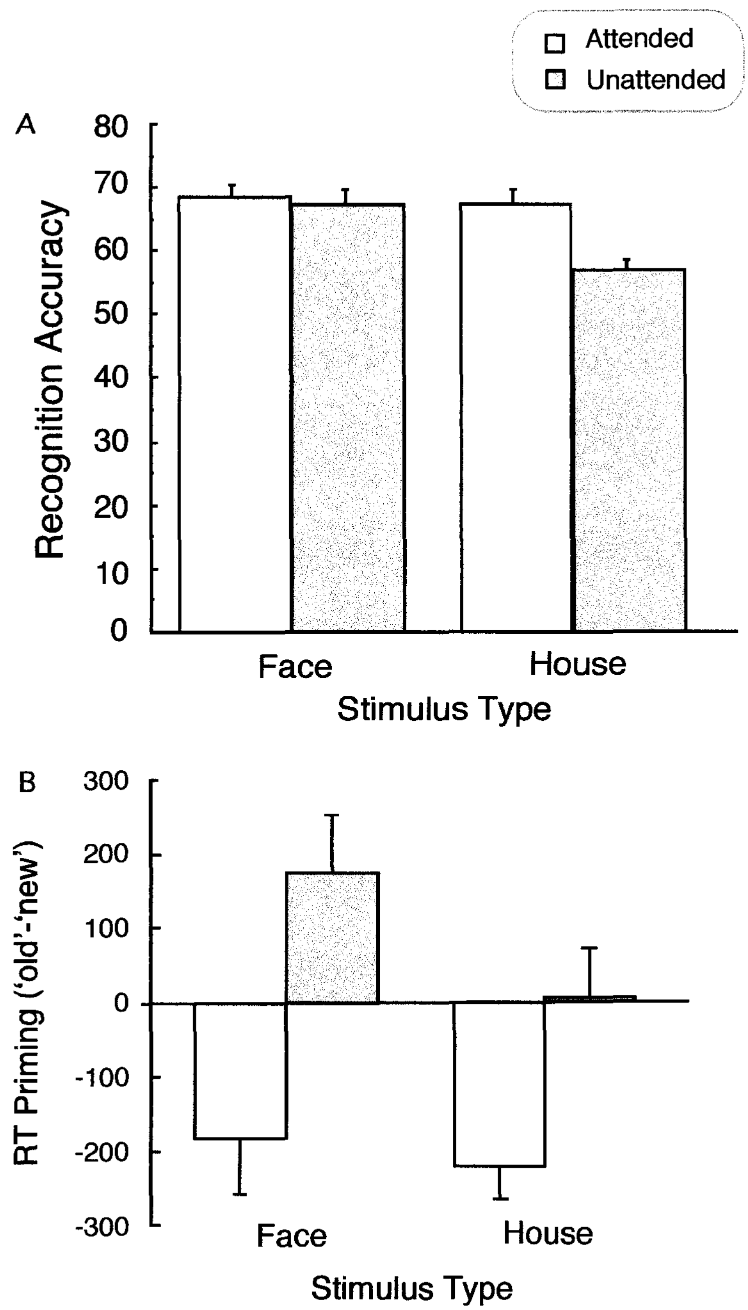

Figure 7. **A** Average face and house recognition accuracy (%) (37 Ss) and **B** Average mean RT priming (msec) for correct responses (37 Ss) on probe obtained in Experiment 3 for attended and unattended conditions. Error Bars represent ± 1 S.E.

IV.6.3 Discussion

The finding that attending to faces interfered with house encoding is indicative of competition between these two image types. Moreover, the finding that face recognition was not affected by attention to the houses suggests an attentional advantage for faces. The RT priming data is consistent with these conclusions. The presence of positive priming for both attended faces and houses, and its absence for unattended stimuli of both categories, suggests that attention had indeed been allocated to the target. The presence of negative priming for unattended faces indicates that faces and objects compete for attention. Negative priming was only found for previously unattended faces and not for previously unattended houses, suggesting that faces were inhibited when attention was allocated to house targets. However, houses that were shown in combination with attended faces were neither inhibited nor activated.

The differential effect of attention on recognition accuracy performance and negative priming we have found may be due to faces drawing more attention than houses. This interpretation is consistent with the results of Vuilleumier (2000), who tested patients with left spatial neglect and found that face stimuli were more likely to overcome extinction than objects and scrambled faces. Mack and Rock (1998) also found that, unlike other stimuli, observers do not fail to detect the presence of happy face icons under conditions of inattention (inattentional blindness). Faces may overcome extinction and be less prone to inattentional blindness because of an advantage in capturing attention. In keeping with this, evidence for an innate advantage in capturing attention can be drawn from studies in developmental psychology. Johnson et al. (1991) found a preference for tracking face-like stimuli in human infants within less than one hour after birth.

If faces draw more attention by default, then their representations would need to be inhibited to prevent them from interfering with the house-matching task. Houghton, Tipper, Weaver, and Shore (1996) found greater negative priming with increased distractor salience. Thus, if distractor faces are attentionally salient, then substantial negative priming should be observed for face distractors. Accordingly, we found a negative priming effect with unattended faces. Furthermore, if faces draw more attention, then there may be little attentional capacity left to be devoted to the irrelevant house stimulus and no need for active inhibition of its representation. This interpretation is consistent with the idea that processing, and thus negative priming of distractors, decreases with increased attentional attraction of attended probes. Lavie and Fox (2000) found that negative priming for distractors decreased with increased perceptual load on target. They concluded that exhausting attention on target processing reduces the amount of processing that the distractor receives, thereby reducing the amount of negative priming. This implies that if faces attract attention and are presented as target stimuli, then very little attentional capacity should be left for the processing of house distractors in our study. Accordingly, house recognition accuracy was significantly influenced when attention was devoted to the faces during presentation of the prime display.

The lack of any priming effect for unattended houses also suggests that either no representation was build for these stimuli or that this representation was not retrieved when attention was devoted to the faces. In contrast, attention to houses did not prevent face representations to be formed as evidenced by the presence of negative priming in the unattended face condition, and by the non-significant difference between recognition of attended and unattended faces. These findings suggest that face representations can be formed with little attention but that house representations cannot. Consequently, the differential recognition accuracy and negative priming effects we have found may be due to both faces being more automatically encoded and faces capturing more attention than non-face objects. It is interesting to note that the results by Vuilleumier (2000) and Mack and Rock (1998) cannot dissociate between these two possibilities. Indeed, faces may overcome extinction and be less prone to inattentional blindness because they are more automatically encoded than non-face objects. Furthermore, Purcell and Steward (1988) showed that masked faces are more easily detected than other masked objects, suggesting highly automatic processing of faces at early processing stages. We therefore conclude that a combined effect of faces being more attentionally salient and requiring less attention to be represented produced the results we obtained.

In conclusion, the results of Experiment 3 confirm and extend the findings of attentional competition between faces and houses of Experiments 1 and 2. In addition, the differential effect on unattended faces and houses indicates that not attending to these stimulus categories seems to trigger differential processes. We suggest that faces are more automatically processed and at the same time attract attention more readily than houses.

IV.7 General Discussion

Although there has been much research on the differential visual analyses performed on face and object stimuli, very little is currently known about the way attention to one stimulus class affects processing of the other. Given our current views on the modular basis of higher visual function, this question remains pivotal because of its implications for the role that higher cognitive factors may play across separate functional processes. In this paper, we have reported the results of three separate experiments that examined whether faces and objects compete for attention and whether their differential processing is reflected in the way attention influences recognition.

The results from Experiments 1A and 2 suggest that simultaneously presented faces and houses compete for attention, irrespective of whether the images are overlapped or spatially separated. The allocation of attention to a given object category produced the best recognition performance when testing for that category. In contrast, directing attention to one object category and testing for recognition of the other yielded poorer recognition performance. The interaction between attention and stimulus type was not significant for both experiments, suggesting that faces and non-face objects were similarly influenced by attention. Finally, recognition accuracy for the unattended images was above chance in these experiments, suggesting that sensory processing alone is sufficient for these images to be represented in memory.

In Experiment 1B, we tested for a face inversion effect with overlapped stimuli. We found that face recognition was more greatly affected by inversion than house recognition. This finding suggests that the faces in the overlapped stimuli were indeed encoded as faces and processed by face-specific mechanisms. The interaction between attention and stimulus type was also not significant in Experiment 1B. Given that inversion eliminates some of the holistic encoding processes that are associated with face processing (Rhodes et al., 1993; Sergent, 1984; Tanaka & Farah, 1993; Tanaka & Sengco, 1997;Young et al., 1987), this finding further suggests that attention exists as a limited resource irrespective of the type of analysis triggered by competing information.

The results of Experiment 3 indicate that attended faces and houses are positively primed. Negative priming was found for unattended faces but not unattended houses. The presence of negative priming for faces is indicative of competition between faces and non-face images. The lack of a negative priming effect for unattended houses shows that attention has a differential effect on encoding of face vs. non-face objects under specific experimental conditions.

Competition between face and non-face stimuli

Attentional modulations similar to those reported here have been previously observed in brain imaging studies (O' Craven et al., 1999; Wojciulik et al., 1998). However, these studies were subject to methodological constraints whereby factors other than competition between face and non-face stimuli could have produced the results (see Introduction). Our findings thus provide the first behavioral evidence for such a competition effect. Furthermore, we showed that this competition is apparent whether attention is allocated in an object– or location–based manner. The finding that the two attentional selection paradigms produced nearly identical results may not be surprising given that both object– and location–based mechanisms are likely to be operative during analysis of a visual scene (Behrmann & Moscovitch, 1994; Humphreys & Riddoch, 1993; Kramer, Weber & Watson, 1997).

We speculate that the same facilitating and inhibiting processes that have been attributed to the attentional selection effects observed in electrophysiological and functional imaging studies may be responsible for our results. For example, it may be that attention to and away from the faces can influence activity in neural networks that are engaged in face processing. One likely candidate for such a function is the fusiform face area (FFA). Indeed, there is some evidence that FFA activity can be modulated by attentional manipulation (Clark et al, 1997; Eimer, 2000; Haxby et al., 1994;). Similarly, attention to and away from the houses may affect neural activity in networks that are recruited by house stimuli. One potential candidate for such an effect is the parahippocampal place area (Aguirre et al., 1998).

The involvement of inhibiting and facilitating mechanisms was more directly investigated in Experiment 3 through use of a priming paradigm. Positive priming is generally taken as evidence for a facilitating mechanism whereas negative priming indicates inhibition (see reviews by Fox, 1995; May et al., 1995 ; Tipper, 2001). We have found that attended faces and houses are positively primed, and that unattended faces are negatively primed. These results further support the notion of a competition between these two stimulus types, whereby selection of one image category facilitates its representation and inhibits that of the other image category.

The lack of a negative priming effect with unattended houses, together with the finding that face recognition was equivalent in the attended and unattended conditions, suggests that competition between faces and non-face objects does not automatically involve inhibition. The visual system must first form a representation of unattended information before it can be inhibited. In addition, if the attended image is easily selected, then inhibition of the competing information may not be necessary (see Stankiewicz, Hummel, & Cooper, 1998 for a similar argument). Our conclusion from Experiment 3 is that such conditions are operative when attention is devoted to a face and recognition of a simultaneously presented unattended house is probed. This conclusion is further discussed in the following section.

Differential influence of attention on faces and non-face objects

Numerous differences in the way faces and non-face stimuli are treated by the visual system have been reported (see reviews by Biederman & Kolacsai,1998 and Tovée, 1998). In the Introduction, we speculated that the differential processing of faces and non-face objects could influence competition between these two categories in four possible ways. One possibility was that faces have an advantage over other stimuli in attracting attention. Another was that faces are more automatically encoded than non-face objects. Evidence for both of these possibilities can be found in the literature (Johnson et al., 1991; Mack & Rock, 1998; Purcell & Stewart, 1988; Vuilleumier, 2000; Vuilleumier & Schwartz, 2001).

A differential effect of attention on faces versus houses was not observed in Experiments 1 and 2. However, such an effect was apparent in Experiment 3. Recognition accuracy data indicates that attention to faces interfered with house recognition but that attention to houses did not interfere with face recognition. In addition, negative priming was found for unattended faces but not unattended houses. We attributed these findings to a combined effect of faces being more automatically encoded and attracting more attention than houses. Specifically, devoting attention to the face by way of a matching task, together with the natural tendency of faces to attract more attention, leaves significantly diminished attentional resources for house processing. The absence of any negative priming for unattended houses suggests that house representations cannot be formed under such conditions. In contrast, the presence of negative priming with the faces suggests that face representations can be formed under conditions of inattention. The negative priming effect also suggests that face representations are inhibited when the house-matching task is performed, perhaps to avoid any interference from the attention-grabbing faces. Thus, faces appear to differ from non-face objects in two ways—first, they attract more attention than non-face objects and second, faces are more automatically encoded than non-face objects such that a face representation can be formed even when attention is devoted to the non-face stimulus category.

We contend that the difference in presentation times between Experiments 1 and 2 versus Experiment 3 represents a crucial parameter. We speculate that faces serve as an exogenous cue to attract attention when they first appear in the visual field, but that a volitional allocation of attention, driven by the task at hand, can overcome this face advantage and favor non-face images. As a result, a tendency for faces to attract attention can be found with paradigms employing short presentation times (less than 1 s) as in Experiment 3 and in the studies on extinction (Vuilleumier, 2000; Vuilleumier & Schwartz, 2001) and inattentional blindness (Mack & Rock, 1998). However, this face advantage is more difficult to detect in paradigms that employ very long presentation times, as in Experiments 1 and 2 (8 s), because there is sufficient time for the initial allocation of attention to the face to be affected by shifts in attention to non-face image tokens. Evidence in favor of a similar distinction between exogenous and volitional attentional mechanisms has been previously reported (Coull, Frith, Buchel & Nobre, 2000; Turatto et al. 2000). Consistent with our interpretation, it appears that exogenous mechanisms are deployed prior to volitional ones (Turatto et al., 2000).

A closer inspection of the data suggests that there is some evidence, albeit not statistically significant, that faces were favored in the allocation of attention in Experiments 1A and 2. Figures 2A and 5 show that the difference between the attended and divided attention condition was more pronounced for house than for face recognition, suggesting that faces may have been favored in the divided attention condition. This is especially evident in Figure 5 where house recognition is almost identical in the divided attention and unattended conditions. It is also interesting to note that in Experiment 1B, where the faces were inverted and presumably did not recruit face-specific processes, the difference between the attended and divided attention condition was similar for inverted faces and houses. We therefore conclude that faces and houses are differentially influenced by attention and that this difference can be detected under specific experimental conditions that involve short presentation times.

Visual search studies indicate that detecting a face amidst face-like distractors (e.g., inverted faces, scrambled faces, sad vs. happy faces) elicits a pattern of RTs that is consistent with serial search (Kuehn & Jolicoeur, 1994; Nothdurft, 1993). These findings suggest that the visual system does not automatically detect facial information at an early stage of visual processing and are inconsistent with our conclusion that faces are more automatically encoded than non-face objects. One explanation as to why faces do not pop-out in visual search experiments may involve the relatively complex nature of the stimulus. Indeed, pop-out is usually found for fairly simple objects, such as geometrical shapes or letters (Treisman & Gelade, 1980; Treisman, 1988; Treisman & Sato, 1990). It is also important to note that in the visual search paradigm, attention may actually be focused in a serial fashion to each stimulus that is present such that there is little competition between the images in the display. Our findings in this study arose from a very different situation where two different object categories were competing for attention. Finally, one must also consider the fact that visual search experiments measure the speed required to detect the presence of a given target among distractors. The attentional face advantage we have found was apparent through an implicit recognition procedure where depth of encoding is presumably more important than speed of detection (see Suzuki & Cavanagh, 1995).

Despite the fact that faces do not pop out in the traditional sense, visual search for faces can nonetheless be faster than for non-face objects. For example, visual search for one's own face is faster than that of a stranger, irrespective of orientation (Tong & Nakayama, 1999). It has also been shown that detection of upright angry facial expressions is faster and more efficient than that for upright happy or inverted angry faces (Fox et al., 2000; see also Oehman, Lundqvist & Esteves, 2001). These findings suggest that faces do hold an advantage in speed of processing even though they do not show the same pop-out qualities as evident with simpler image tokens.

The problem of stimulus equivalence

Physical stimulus features are always difficult to control whenever different classes of stimuli are compared. The choice of a stimulus category that is comparable to faces is particularly problematic given their idiosyncratic properties. Houses have been used in previous studies and in the present one because they share some of the same general characteristics as faces, i.e., as with faces, all houses share a set of common features that are arranged in a specific manner that renders each exemplar unique. However, surface properties such as spatial frequency information are difficult to equate. Moreover, the high level of homogeneity between different individual faces can hardly be matched even with careful selection of comparison stimuli.

A more pervasive aspect of face perception that is rarely found for other object categories is expertise. This important characteristic is believed to account for many of the functional and anatomical processing differences between these two stimulus classes (Diamond & Carey, 1986; Gauthier & Tarr, 1997; Gauthier et al., 1998). Indeed, some researchers argue that face processing is not "special" but rather exemplifies the workings of a visual recognition system responsible for subordinate-level classification of visual stimuli for which we are experts. In support of the expertise hypothesis, many face-specific effects have been observed for non-face objects among relevant experts. It has also been shown that non-face objects for which expertise has been acquired can produce activation in a region of the human fusiform gyrus that is preferentially activated by faces (Gauthier et al., 1999; 2000).

The evidence in favor of the expertise hypothesis, however, does not preclude the existence of specific mechanisms for face recognition. Even if faces are in fact processed by a system devoted to discriminating any visually homogeneous stimulus, then it may be the case that for most people faces are the only object category that meets the sufficient, but not necessary, criteria to recruit this system. Faces therefore constitute ideal stimuli for investigating the underpinnings of a specialized functional module irrespective of whether the domain of specialization of this module is face perception *per se* or expertise for within-category discriminations.

Modules and attention

The evidence in favor of a dedicated anatomical region for face processing, together with the numerous functional differences that exist between face and non-face stimuli, point to the existence of a specialized functional module for face perception. Whereas the characteristics of a modular cognitive architecture have been well defined (Fodor, 1983), to what extent modules are autonomous is less understood. Our finding that faces and non-face objects compete for attention may have implications for this issue. Indeed, this competition implies that attention is not module-specific but rather exists as a generalized resource that is allocated to multiple domains of visual processing. This notion is consistent with Fodor's speculation that mental processes often compete for access to resources, such as attention or short-term memory. Allocating these resources to one of the competing processes produces a decrement in performance in the others.

Fodor (1983) also states that information processing by a module is mandatory such that the operations of any given module are automatically triggered in the presence of its dedicated stimulus. One important consequence of this type of analysis is speed. Our results suggest that faces are more automatically encoded and attract more attention than non-face objects under short presentation times. This conclusion implies that the presence of a face is detected early in the analysis of a visual scene. It may be that the fast and relatively automatic analysis of faces arises because they are processed by a specialized cognitive module.

Conclusion

We have shown that faces and non-face objects compete for attention. If faces are indeed processed by a specialized functional module, this finding suggests that such modular processes are subject to the same attentional limitations as other visual operations. We have also found evidence that faces and non-face objects are differentially influenced by attention, with faces attracting more attention and being more automatically encoded than non-face objects. This difference adds to the list of previously reported functional differences between faces and other complex stimuli and as such further supports the notion that faces engage unique processing mechanisms. We suggest that this effect is only present at short presentation times because faces serve as an exogenous cue for attention in the early stages of visual analysis. Whether or not this attentional advantage is the result of faces being processed by a dedicated cognitive module, and as such applies to modular processing in other domains as well, remains to be determined.

General Discussion

Chapter 5

Discussion

The important status that faces hold in conveying personal and emotional information, together with the difficulties inherent in recognizing faces, has led researchers to suggest that faces are processed by a specialized cognitive module of the type postulated by Fodor (1983). Cognitive modules are impenetrable input analyzers that are fully devoted to a specific function and that are activated in an automatic and obligatory manner by stimuli belonging to their proprietary domain. Although different modules have different specializations, they nonetheless possess generalizable properties that are common to all modules. As such, the modularity hypothesis provides a constructive theoretical foundation for the investigation of a broad range of psychological functions that have properties that may be applicable across different cognitive domains.

Efforts to link modularity with face recognition have mainly focused on the issue of 'specialness' with numerous attempts to pinpoint functional and anatomical differences between face and object processing. These investigations have generally been successful in showing that faces engage operations that are specialized for their inherent characteristics, and in identifying a cortical region that is likely responsible for face processing. Some progress has also been made in deciphering the contribution of genetic and environmental factors in the development of face recognition abilities. Hence, it is generally agreed that face recognition is modular because the operations underlying this function are domain specific, innately specified, and subserved by a fixed neuronal architecture that displays specific breakdown patterns.

Despite an extensive body of literature on faces, very few studies have examined whether other equally important aspects of modularity are associated with face perception. These include informational encapsulation, limited access to central processes, mandatory processing, fast operations, and shallow outputs. This thesis strengthens evidence that face perception may indeed be classified as modular by linking some of these properties with face processing. Starting from the assumption that faces exemplify modular functions, the thesis also examines aspects of modularity that are less well understood, with a special emphasis on the role that attention plays in a modular cognitive architecture.

5.1 Mandatory processing of faces

Modular operations are mandatory in that they are automatically applied to stimuli in the module's proprietary domain. In Chapter 2, mandatory face processing was demonstrated using the composite effect, a well-established test of holistic face encoding. This is important because in order to provide evidence that processing within the face module is automatic, one needs to show that operations specific to that module can be carried out without attention. Previous research as well as the results presented here (Chapter 3) point to a distinction between encoding of faces as Gestalts and encoding of non-face objects as features. The results indicate that diverting attention away from faces by way of two different manipulations during encoding did not eliminate holistic encoding, as measured by the composite effect. This finding provides a significant contribution to the debate on modularity in that it links a previously ignored modular property to face perception.

The results presented in Chapter 2 represent the first evidence that specialized operations are automatically applied to faces. Previous research by Reinitz and colleagues failed to produce consistent results (Reinitz et al. 1994; 1997). Moreover, these studies were limited because they employed line-drawn stimuli that may not have optimally triggered holistic face encoding processes (Leder, 1996). More convincing is evidence put forward by Palermo and Rhodes

(2001). They have shown that matching two faces presented on each side of a target face during learning eliminates holistic encoding of the target face as measured by the whole/part advantage at testing.

Two factors may account for the discrepancy between the results presented in Chapter 2 and those of Palermo and Rhodes. First, a different procedure was employed for assessing holistic encoding. In Chapter 2, the composite effect was used where the bottom half of one face interferes with recognition of the top half of another face. Palermo and Rhodes measured holistic face encoding using the whole/part advantage where recognition of a face part is superior in the context of a face than in isolation.

It is not clear which of the two procedures provides a better assessment of face-specific operations. Whereas attempts have been made to provide precise definitions for the terms holistic and configural, these two constructs are often used interchangeably in the literature and different operational definitions have provided conflicting results (e.g., Leder & Bruce, 2000; Tanaka & Farah, 1993; Tanaka & Sengco, 1997). The composite effect arguably taps into both configural and holistic information processing (Diamond & Carey, 1994). In aligned composites, the features in the top half of the learned face are arranged in a new spatial arrangement with the features of the bottom half of another face. In addition, presentation of old features in the context of a new face disrupts the holistic representation stored at learning. However, others have suggested that composites may reflect holistic encoding processes rather than configural ones (Hole et al., 1999). In contrast, the whole/part advantage used by Palermo and Rhodes (2001) is an unambiguous test of holistic encoding. Furthermore, this effect has the advantage of having been directly compared across faces and non-face objects (Farah & Tanaka, 1993). There are no reports that the composite effect does not apply to non-face objects. However, it is generally assumed that this phenomenon is restricted to faces. This issue will remain unresolved until demonstrations to the contrary are provided.

One important difference between the experiments presented in Chapter 2 and those by Palermo and Rhodes (2001) rests in the attentional manipulation used. The manipulation used in Experiment 1 was unsatisfactory because the effect of attention and low-level segregation mechanisms cannot be dissociated when overlapped transparent stimuli are employed. However, the manipulation used in Experiment 2 was not limited by this confound. A distractor task similar to those employed in previous studies on attention was used (e.g., Reinitz et al, 1994; Eimer, 2000). The manipulation used by Palermo and Rhodes is problematic because it required that two flanker faces be matched during the presentation of the target face. Matching of the two flanker faces probably required that each face be holistically encoded. As a result, all three faces presented during encoding would have had to be holistically encoded in order for the unattended face to produce a whole/part advantage. Therefore, the results presented by Palermo and Rhodes can be attributed to a limitation in the number of faces that one can holistically encode at any given time, rather than to a failure for faces to automatically trigger holistic operations. Another difference rests in the dependent variable employed in the two studies. Whereas the results presented here are based on a d-prime measure of sensitivity, those by Palermo and Rhodes (2001) are based on recognition accuracy. Unattended faces may have produced a significant difference between recognition of face parts presented in isolation and those presented in the context of a face if a more sensitive measure was used.

The discrepancy between the results presented here and those of Palermo and Rhodes (2001) warrants further experimentation. The issue of mandatory face processing would best be resolved using a variety of techniques designed to assess holistic face encoding and other face-specific operations. Future investigations should carefully consider which attentional manipulation and recognition performance measure is more appropriate. All studies on attention suffer from the same limitation in that they can be criticized on the grounds that attention can never be entirely removed from a given stimulus and that even a little attention may be sufficient to replicate the effects found in a full attention condition. The attentional manipulation used here was clearly efficient because it significantly reduced recognition of faces. What is important is that distracting attention interfered with the *degree* that faces were encoded but not with the *way* that they were encoded. This finding provides an initial demonstration that face-specific operations meet the automaticity criterion of modularity.

5.2 A rivalry phenomenon that supports modularity of face processing

Chapter 3 investigated possible determinants of the rivalry effect produced by overlapping upright tilted faces. Although these experiments focused mainly on the relationship between grouping principles and competition for visual awareness, they also help us to broaden our understanding of face perception. Perceptual rivalry was evaluated on the basis of the participants' ability to clearly perceive and recognize both faces in the stimuli for three different presentation durations. It was reasoned that rivalry would prevent perception and recognition of both tilted images in the overlapped displays at short presentation times, but not at longer times where an alternation from one tilted image to the next could occur. This is exactly what was found with overlapped upright faces irrespective of whether they were photographic positive or negative. For the overlapped inverted faces and overlapped houses, only one tilted image was clearly perceived or accurately recognized at all presentation times, suggesting that these displays do not readily produce rivalry.

Because rivalry was measured as a function of the perception and recognition of either one or both of the tilted images over time, two possible explanations can be offered to explain the results. It may be that non-face images did not produce rivalry within the inspection time allowed because it took longer for these images to be perceived as Gestalts. Alternatively, non-face images may not produce rivalry because they did not trigger Gestalt encoding processes at all. With respect to the modularity hypothesis, the former possibility would indicate that faces are processed more quickly than other objects. The latter would suggest that faces engage domain specific operations.

The favored interpretation for the results is that specialized holistic operations are applied to faces. Fodor (1983) asserts that modules are domain specific because the stimuli they process have idiosyncratic properties that require tailored operations. Accordingly, it has been argued that identification of highly homogeneous faces at the individual level (e.g., John's face) necessitates the use of configural or holistic information (Carey & Diamond, 1994; Farah et al., 1995a; Yin, 1969). In contrast, piecemeal information may be sufficient for recognition of non-face objects at the categorical level (e.g., car, chair, etc.). The rivalry effect observed for overlapped faces, and the lack thereof with inverted faces and non-face objects, suggests that faces do trigger holistic encoding operations that are closely tuned to their idiosyncratic properties. This demonstration of domain specificity has the advantage that the same manipulation was applied to faces and non-face objects. Indeed, only a few studies have directly compared faces with other complex objects (Gauthier et al., 1998; Tanaka & Farah, 1993; Tanaka & Sengco, 1997) and most of the evidence cited to support the importance of holistic information in face recognition rests on the face inversion effect (Leder & Bruce, 2000; Rhodes et al., 1993; Searcy & Bartlett, 1996; Sergent, 1984; Young et al., 1987).

The possibility that overlapped non-face objects did not produce rivalry because they are processed more slowly than faces needs further consideration. Inspection of the stimuli (see Figure 1, Manuscript III) demonstrates that one upright face can be quickly segregated from the other without voluntary effort. In contrast, segregation of one house from the other is timely and difficult. This suggests that the rivalry effect is due to a combination of faces triggering Gestalt encoding and the speed with which this process takes place. This would mean that faces are not only processed by specialized operations but also that these operations are fast. Both properties are compatible with the modularity hypothesis. Although the rivalry effect provides indirect evidence for fast face-specific operations, more direct evidence from electrophysiological studies in monkeys and ERP (event related potential) studies in humans corroborate this contention (Jeffreys, 1996; Wallis & Rolls, 1997).

The finding that one of the faces in the overlapped display was quickly and automatically perceived as a Gestalt further suggests that holistic face encoding is mandatory, as suggested in Chapter 2. The rivalry effect also suggests that these operations cannot be applied to two faces simultaneously. According to Fodor (1983), modular operations are mandatory and fast because of informational encapsulation. Informational encapsulation refers to the notion that modular outputs cannot be modified to match our beliefs or desires. Visual illusions such as the Ames room illustrate this property. Whereas illusions may indicate that modules are 'dumb' in that their outputs cannot be modified by beliefs in erroneous situations, they also show that only predetermined ecologically plausible conditions are tacitly present in modular processes. Modules 'know' about the most probable structure of a stimulus and it is because of this knowledge that their operations are fast and efficient.

Multistable percepts have traditionally been attributed to some implicit awareness that two things cannot occupy the same place at the same time in the natural world. This raises the interesting possibility that the rivalry effect is, like other illusions, a direct consequence of informational encapsulation. This would in turn be consistent with the notion that faces trigger mandatory and fast operations, both of which are derived from informational encapsulation. Although the notion that rivalry arises from informational encapsulation is speculative, our everyday experience suggests that humans do in fact have very little voluntary control over face recognition. It does not seem possible to instruct oneself not to recognize the familiarity of the faces we encounter, or not to perceive a face in a Gestalt fashion.

The finding that rivalry was influenced by inversion, but not contrast reversal, points to an important limitation in our understanding of face-specific operations. The interpretation offered for this finding is that these two manipulations tap into different aspects of face encoding, with contrast reversal disrupting configural processes and inversion holistic ones. However, the terms configural and holistic are often used interchangeably in the literature and the use of different methodologies to evaluate these constructs has often led to inconsistent results (e.g., Leder & Bruce, 2000; Tanaka & Sengco, 1997). Some researchers have suggested that holistic encoding is initially applied to faces as a way of determining that a face is a face. Consideration of configural information would take place in a subsequent stage to allow for identification of who's face it is (Carey & Diamond, 1994; Hole et al., 1999). This distinction in terms of time of processing may prove useful in future investigations, especially with high temporal resolution imaging techniques that may be sufficiently precise to dissociate the operations underlying these two encoding strategies.

Attempts to link modular properties with the rivalry effect produced by overlapped faces are limited by the fact that such stimuli are not ecologically valid. Nonetheless, investigating this effect has provided a compelling illustration of how modular processes may operate under such conditions. Additional contributing factors should also be explored, perhaps by testing for the influence of various stimulus transformations on the rivalry effect. Showing that rivalry fails with other non-face categories would also be useful in providing further support for domain specificity of face perception.

5.3 Beyond modular properties of face perception

Accepting the modularity hypothesis of face perception offers the interesting possibility of using faces to investigate aspects of a modular cognitive architecture that are less well understood. This approach was adopted in Chapter 4 where the interaction between attention and modular operations was examined. Two specific questions were addressed. First, whether attention is domain-specific or domain-general was investigated. Second, whether modular processing confers an attentional advantage to faces was examined.

5.3.1 Domain-specific versus domain-general accounts of attention

In his discussion of informational encapsulation, Fodor (1983) states:

> [...] a picture that is now widely accepted among cognitive psychologists: Mental processes often compete for access to resources variously characterized as attention, short-term memory, or work space; and the result of allocating such resources to one of the competing processes is a decrement in the performance of the others. (p.72)

This statement raises an important question: do modules have dedicated attentional resources (domain-specific) or do they share a common pool (domain-general)? Whereas Fodor's statement favors a module-general account of attention, results from two studies suggest that the module-specific account may apply to the visual recognition system. First, Khurana (2000) found that devoting attention to distractor upside-down faces during the presentation of a probe target face does not influence the priming effect produced by the target. Second, Palermo and Rhodes (2001) found that matching inverted flanker faces presented on each side of an upright target face did not significantly impair holistic encoding of the target. Both studies indicate that devoting attention to inverted faces does not interfere with encoding of an upright face. It has been argued that inverted faces do not trigger the same holistic encoding operations as upright faces but rather are processed by a separate general object recognition module (Farah et al., 1995b; Moscovitch et al., 1997). If this is the case, then the finding that inverted and upright faces do not compete for attention would support a domain-specific account of attention.

Studies that have examined the influence of inversion on responses in the putative face area, the FFA, indicate that this manipulation diminishes the area's selectivity for faces but does not eliminate it (Haxby et al., 1999; Kanwisher et al., 1998). In addition, inverting faces increases activation in areas that are usually activated by non-face objects, suggesting that inverted faces can be processed by both face and non-face areas. The finding that prosopagnosia does not impair recognition of inverted faces but in fact improves it further supports the notion that inverted faces can be processed by areas that are not involved in face recognition (Farah, et al., 1995b). Finally, Moscovitch and Moscovitch (2000) have proposed a model of face recognition where the object system first analyses inverted faces and then provides this information to the face system. Consistent with this model is the finding that the FFA activation produced by inverted faces is comparable to that produced by upright faces but with a slight delay (D'Esposito, Zarahan & Aguirre, 1999). Therefore, inverted and upright faces may not compete for attention because they are processed by different areas, as predicted by the domain-specific account of attention.

In Chapter 4, attentional competition between faces and houses was investigated using three different paradigms, all producing essentially the same result. Devoting attention to a face by way of a behavioral task significantly reduced recognition of a simultaneously presented house, and vice-versa. This finding is consistent with a domain-general account of attention whereby attention is a generalized resource that has extensive connections with various modular processes. The results are also consistent with a generally accepted notion that devoting resources to a given module limits other modules' access to these resources.

The results presented in Chapter 4 contradict those reported by Khurana (2000) and Palermo and Rhodes (2001). This discrepancy may be related to their use of inverted faces and the use of non-face objects in the studies presented here. The fact that inverted faces activate the FFA suggests that they may also trigger modular face-specific operations. As such, processing of inverted faces may be partially mediated by mandatory processes of the type postulated for upright faces. Therefore, analysis of inverted faces may not interfere with analysis of upright faces because few attentional resources are used for both processes. In contrast, houses presumably do not trigger operations carried out by the face module and as such, house analysis may be more taxing for the attentional system. This possibility is supported by the interference effect reported here.

The proposal that inverted faces may recruit mandatory operations because of their representation in both a face and an object system implies a greater interference of distractor houses on target upright faces than of distractor inverted faces on target upright faces. Further experimentation is needed to determine whether this is the case. If the discrepancy between the results obtained here and those reported elsewhere (Khurana, 2000; Palermo & Rhodes, 2001) does rest in the propensity for inverted faces to trigger mandatory operations carried out by the face module, then inverted faces may not be ideal for investigating the issue of whether modules have their own dedicated resources or share a common pool.

5.3.2 Do faces hold an advantage in the allocation of attention?

A second purpose of the experiments presented in Chapter 4 was to examine whether faces hold an advantage in the allocation of attention. A number of studies suggest that this may be the case. Of particular relevance are the findings that the emotional content of faces can be detected pre-attentively in a visual search task (Öhman et al., 2001; White, 1995; but see Fox et al., 2000; Purcel et al., 1996), that happy face icons can be detected despite inattention (Mack & Rock, 1998), and that neutral faces are more easily detected than other types of images in the 'attentionally blind' side of patients suffering from spatial neglect (Vuilleumier, 2000; Vuilleumier & Schwartz, 2001). In Chapter 4 (Experiment 3), corroborating evidence for the notion that faces hold a special status in the allocation of attention was provided. It was found that with short presentation times, faces are favoured in the allocation of attention. Two interacting factors were proposed to account for this finding —(i) faces may capture attention more readily, and (ii) they may be more automatically encoded than other objects.

The proposed tendency for faces to capture attention may arise from the fact that certain sub-cortical and parietal brain areas are involved in face processing. Haxby and colleagues (Haxby, Hoffman & Gobbini, 2000) proposed a model that states that changeable aspects of faces i.e., expression, eye gaze, and lip movements, are analyzed by a common neuronal system that is independent from that underlying recognition of identity. The model is based on imaging and electophysiological data as well as neuropsychological evidence that disorders in recognizing changeable aspects of faces can be independent from disorders in recognizing facial identity. The model also includes extensive connections between the multiple regions involved in face perception. As a result processing a given face attribute may enhance responses in the region primarily responsible for analyzing that attribute as well as in regions activated by other facial attributes.

Brain regions that are activated by changing aspects of a face and that may mediate attentional shifts towards faces include the amygdala, the pulvinar, and the parietal cortex. Faces with a strong expression i.e., angry or happy faces, can modulate activation in the amygdala even if they are not consciously perceived (Whalen et al., 1998) or attended to (Vuilleumier et al., 2001). Conditioned emotional faces have also been shown to produce activation in the pulvinar (Morris, Firston & Dolan, 1997). Perception of averted gaze produces activity in parietal regions that are responsible for shifts in spatial attention even if gaze direction is irrelevant to performing the task at hand (Langton & Bruce, 1999). This suggests that gaze perception may direct attentional shifts towards faces in the very early stages of face processing. Finally, there is evidence that emotions and face configuration can be detected before the level at which attentional mechanisms operate (Vuilleumier, 2000; Vuilleumier & Schwartz, 2001). These studies suggest that increased sub-cortical and parietal responses to faces may subsequently enhance responses in face processing areas that process facial identity. In keeping with this, it has been suggested that humans may be born with a special mechanism that automatically brings their attention to faces (Morton and Johnson, 1991).

One limitation in integrating previous findings from the literature on expression and gaze perception with the results reported here is that the faces employed in the present studies had neutral expressions and a viewer-oriented gaze. With respect to gaze, there is some evidence that parietal regions are more activated by faces than by other objects irrespective of gaze direction (Clark et al., 1998). Given that parietal region are involved in shifts in attention, this finding suggests that neutral faces may also capture attention via activation in parietal regions. With respect to emotions, it has been shown that detection of a familiar neutral face produces greater amygdala activation than for discrimination of the direction that the face was rotated (Sugiura et al, 2001). Identification of expression in neutral faces also appears to produce activation in the amygdala. Therefore, the proposed capturing of attention by neutral faces may be mediated by amygdala activation.

The finding that faces maintain an advantage for the allocation of attention may therefore be attributed to the concerted action of the different brain areas involved in processing different aspects of faces. In this speculative model, analysis of gaze and emotions would initiate shifts of attention to faces and in turn facilitate identification of a face as a face. Once attention is directed to a face and it has been recognized as such, determining whether the face is familiar may be accomplished via the three-stage model proposed by Bruce and Young (1986). In the Bruce and Young model, recognition of a familiar face begins with activation of a face recognition unit that attempts to match the face's stored description. If the level of activation reaches a certain threshold, the face is recognized as familiar. It is during this stage that faces would benefit from being processed by a dedicated module whose operations are mandatory. Matching a holistic representation of a viewed face with its stored equivalent appears to be automatic. The results presented in Chapter 2 and 4 are consistent with this proposal. The initial capture of attention by faces may be attributable to their social significance.

The interpretation offered for the results presented in Chapter 4 may be limited because recognition performance was used to make inferences about processes that take place during encoding. Indeed, the interference effects and attentional face advantage observed in Chapter 4 may arise from either the way that faces are analyzed during encoding, or from the way that their representations are stored in memory. In fact, this criticism applies to most of the data presented. It would be important for future research to replicate these findings using immediate perception paradigms such as reaction time measures and simultaneous matching rather than recognition procedures. Imaging techniques could also be useful in determining whether activation in brain regions that mediate shifts in attention can modulate subsequent FFA responses to faces. Finally, the interaction between attention and modular functions should be further investigated to determine whether the attentional advantage observed here and elsewhere (Mack & Rock, 1998; Purcell & Stewart, 1988; Vuilleumier, 2000) is unique to faces or applies equally well to other potentially modular functions.

5.4 Modularity and face recognition models

Several face recognition models have been proposed over the last twenty years and the conceptualization of face processing as modularly organized has important implications for these models. Face recognition models may be divided into several different categories. The first distinction is between those models that can fit under the umbrella of modularity and those that rely on a distributed account. The modular account has been thoroughly defined by now and the different theories that fall under that category are further described below. The distributed account has been touched on in the discussion of the Haxby et al. (2000) model. In distributed models, many different types of information are processed by many different parts of the brain such that any one brain region is likely to represent many different classes of stimuli. In the Haxby et al. (2000) model, perception of different aspects of faces is conceived of as arising from activation in a widely distributed network of sub-cortical and cortical areas.

A more recent model by the same group (Haxby et al., 2001) stipulates that widely distributed and overlapping networks in the ventral temporal cortex mediate perception of both faces and objects. Whereas this model provides an interesting alternative to the modularity account, it is difficult to reconcile the neuroimaging data provided to support the model with the existence of a double dissociation between prosopagnosia and agnosia for other object categories (Moscovitch et al., 1997). Furthermore, explanations based on modularity may equally well fit the data if not better. For example, Haxby et al. report that faces produce activation outside of the FFA. However, this activation may not contribute to the conscious experience of face perception and recognition.

Another possibility is that the visual recognition system is organized hierarchically with face selective areas representing a superior level of analysis that uses information from lower levels to process face information (Moscovitch & Moscovitch, 2000). The modular account therefore appears more plausible than the distributed account on the basis of the evidence currently available. Nevertheless, a reassessment of the modularity account may be necessary if future investigations were to produce results in favor of the model proposed by Haxby and colleagues.

Modular accounts of face perception may be further distinguished on the basis of domain specificity. The debate here rests on whether the face module is specialized for face recognition *per se* or for recognition of any object class for which expertise has been developed. Gauthier and colleagues (Gauthier & Tarr, 1997; Gauthier et al., 1998; Gauthier, et al., 1999a; Gauthier et al., 1999b; Gauthier et al., 2000) have been at the forefront of this debate by providing compelling evidence that FFA activation can be modulated by expertise for non-face objects, and that presumed face specific effects can apply to non-face objects in experts (see Section 1.3.1.6 of the Introduction for a more complete description of this argument). Therefore, expertise and homogeneity may be responsible for the holistic operations involved in face perception and recognition. Hence, the face module may best be characterized as a system that codes relational features. As Kanwisher (2000) pointed out, whether the issue of expertise poses a serious obstacle to modular accounts of face perception really depends on one's conception of domain specificity.

In the thesis, the distinction between featural encoding of objects and holistic encoding of faces is taken as evidence for domain specificity. The thesis also stresses the quantitative nature of this distinction. However, it may be argued that qualitative differences are necessary for the attribution of domain specificity. Kanwisher (2000) states many different criteria from which domain specificity can be evaluated. In her opinion, face processing is likely to be domain specific because it meets the criteria of common use (faces are the most common use for the module), necessity (the module is necessary for face recognition), and origins (innateness of the face module). The same argument is made here. It seems reasonable to argue that these criteria are sufficient in the case of faces even though other modules may meet different criteria.

The expertise hypothesis has implications for what Kanwishser (2000) refers to as the criteria of 'possible use' of a module. As such, the face module may sometimes by used for processing of other object categories if the category necessitates recognition at the individual level and if expertise in doing so has been acquired. However, this possibility does not preclude the use of faces for investigating a modular cognitive architecture. Furthermore, the criterion of 'possible use' may be a weak indicator of domain specificity (Kanwisher, 2000). Perhaps even more important is the fact that the expertise hypothesis does not contradict the notion of an innate module that is dedicated for face perception. It simply suggests that this module may also be recruited by other tasks under very specific conditions that involve extensive training and discrimination of highly homogeneous stimuli on the basis of holistic information.

Numerous face recognition models explicitly or implicitly rely on the notion that faces are processed by a domain specific modular system. Those may be distinguished at different levels. Models that focus on the type of information used to recognize faces can be classified as relying on holistic, configural, norm-based, and exemplar-based representations. The first two accounts have been extensively reviewed in various sections of the thesis. The norm-based model encompasses various hypotheses, all of which assume that storing representations of faces in memory entails the abstraction of something that can be called a face norm, prototype, or schema. In exemplar-based models, there is no extraction of a norm or prototype from a face but rather faces are considered as being encoded as points (reviewed in Valentine, 1991).

A second possible classification for modular models of face recognition arises from the nature of the different stages that are responsible for this function. A prevailing model in that regard is the one put forth by Bruce and Young (1986). In the model, faces are represented on the basis of a *structural code* that is essential for distinguishing one face from another. *Structural codes* are view-centered and different types may exist for familiar and unfamiliar faces. Familiar faces are associated with *identity-specific semantic codes* that encompass various information associated with the identity of the familiar person such as employment, location of where the person is usually encountered, etc. Familiar faces are also associated with an independent *name code*. Unfamiliar faces are associated with *visually derived semantic codes* that are largely based on appearance and that encompass attributions of character and resemblance with known individuals. *Expression codes* and *facial speech codes* (lip movements) may also be created upon perception of both familiar and unfamiliar faces.

In the Bruce and Young model, these different codes specify the functional components of the human face processing system. In the first stage of face processing, *structural, expression* and *facial speech codes* are derived from a viewed face. Familiarity is determined by matching the *structural code* with an equivalent code stored in memory. These stored codes are called *face recognition units*. *Face recognition units* send a signal to the rest of the cognitive system irrespective of whether the face is familiar or not. It is the strength of this signal that will determine familiarity. In the second stage, activation of a *face recognition unit* provides access to stored semantic information about the familiar person. This information is held in a portion of associative memory called *person identity nodes*. *Name codes* are retrieved in a final stage via the person identity nodes.

Although the Bruce and Young (1986) model focuses on the functional components of the face processing system, converging evidence from lesion studies suggest that anatomically distinct brain areas may in fact mediate the proposed functions. These include double dissociations for disorders of familiar and unfamiliar face recognition, disorders of face recognition and analysis of facial emotions, and disorders of facial speech and other aspects of face processing (reviewed by Young, Newcombe, de Haan, Small and Hay, 1993).

Existing face recognition models such as the one by Bruce and Young should be improved or modified to encompass the modular characteristics suggested here. First, Chapter 2 indicates that faces trigger mandatory holistic processing operations. This finding could be incorporated in the Bruce and Young model in such a way that perception of a face would have to automatically activate relevant *face recognition units*. Second, the results presented in Chapter 4 favor a modified model that combines face recognition stages similar to those proposed by Bruce and Young together with mechanisms that direct attention to faces in the early stages of visual processing. Finally, mediating mechanisms that are not currently considered in several models of face recognition should also be taken into account. Of particular relevance to the present discussion is the issue of visual awareness. Covert recognition of faces has been demonstrated in some prosopagnosic patients using both behavioral and physiological studies. In view of this evidence, De Haan (2001) suggests that the different processing stages involved in face recognition are largely automatic and escape conscious awareness. This interpretation, together with the observed double dissociations between different aspects of face perception, raises the possibility that the face module is divided into different sub-modules, each producing a cognitively impenetrable representation that is later integrated with what is produced by other sub-modules. Future models should discuss the idea that only the output of these integrated processes may be available for conscious report, as suggested by Fodor (1983). Such models may provide a more appropriate framework for the interpretation of the rivalry effect observed here as well as to other illusions of face perception.

5.5 Why is modularity a useful hypothesis?

The aim of the thesis was to link data obtained in the field of face perception with the modularity theory proposed by Fodor (1983). The contribution of such an effort rests in the belief that modularity is a useful foundation for research in face perception and in cognitive science in general. Both inductive and deductive approaches to science can serve to refine our understanding of normal cognition. The former approach has the advantage that specific questions may be directly addressed via a wide variety of available empirical methods. This is exactly what Fodor hoped to accomplish by proposing a theory of modularity. He wrote:

> I am not, in any strict sense, in the business of defining my terms [...]. So what I propose to do instead of defining modularity is to associate the notion of modularity with a pattern of answers to such question as 1-5 (p. 37).

Here Fodor refers to five possible modular characteristics. He later stresses that cognitive scientists should focus on finding potentially modular functions and on determining which modular properties are associated with that function. He believed that to discover and characterize the fundamental components of the mind, one has to seek to identify and further characterize modular systems. In fact, Fodor (1983) believes that cognitive modules are the only valid topic of scientific inquiry because only modular processes are sufficiently local to have properties that are common over and above properties that define a particular cognitive function. In contrast, central processes lie outside of our investigative reach because there is no specific experience that can decisively confirm or contradict a belief, and because beliefs are subject to our individual histories and exigencies of the moment. Whereas this proposition may be perceived as a gloomy prognosis, it offers an important contribution to psychology. By postulating a set of defined characteristics that can be directly tested across a wide variety of systems, Fodor has led the way for the development of a psychological theory that may be applicable to several domains. Adopting this approach has proven fruitful here as in other scientific endeavors in that it allows for a global understanding of cognitive functions as well as for the discovery of new and unexpected findings that may lead to further experimentation.

It is surprising that while modularity is generally taken for granted in psychology, it is often misunderstood. One common misconception is that all modular characteristics proposed by Fodor (1983) must be met for the system to be classified as modular. However, Fodor notes that the modular properties he described are not conditions that have to be met in order for a cognitive function to be classified as modular. Rather, a number of these conditions must be true for the attribution of modularity. Fodor also suggested that if several modular properties are true for a particular cognitive function, then most of them are likely to be realized. The case is made here that the modularity thesis is most valuable when all the properties that are associated with a given function are considered. By doing so, one can provide a basis from which other potentially modular functions can be examined.

Other misconceptions relate to which of the properties are more important for modularity. It is often believed that domain specificity is the most essential condition for modularity but in fact Fodor finds informational encapsulation to be more defining (Coltheart, 1999). This further stresses the importance of exploring all aspects of modularity. Finally, misconceptions with respect to the exact meaning of the different properties are also common. For example, domain specificity does not only imply that modules analyze a particular stimulus class but also that the module carries out particular operations that are tailored to the idiosyncratic properties of that class. This distinction between class and function specificity is not clearly made in Fodor's treatise and as previously discussed, both may be appropriate definitions of domain specificity. Another relevant example relates to automaticity. Unlike previously suggested (Wojciulik et al., 1998), finding that activation in the FFA is modulated by attention does not invalidate the notion of mandatory processing in face perception. To do so, one would have to show that distracting attention from faces somehow modifies the *way* in which faces are encoded.

The modularity hypothesis was used in this thesis as a way to broaden our knowledge of face perception and to provide a foundation from which to investigate other psychological functions that may be classified as modular. Using such a unified account to examine different psychological functions has proven especially valuable in the case of audition where a modular system similar to that which has been proposed for faces may exist. Indeed, recent evidence suggests that the visual and auditory systems are analogous in that voices selectively activate a discrete region in the superior temporal sulcus (Belin, Zatorre, Lafaille, Ahad & Pike, 2000). Like faces, voices can be used to identify individuals. They also play an important role in social interactions where voice-related information such as emotion is crucial. This finding raises the interesting and perhaps not surprising possibility that socially relevant information is encoded by specialized modular systems. After all, modules are supposed to allow for a quick and efficient analysis of the input. This feature may be particularly helpful in the case of information that has high social value such as affect and identity. Linking modular characteristics with face perception can thus serve for later comparison with other potentially modular functions. It is important to mention that empirical findings that are taken as evidence for a modular cognitive architecture may also be consistent with the main alternative, the distributed account of cognition (e.g., Grossberg, 2000; Zeki, 1998). As such, the modular theory is not useful because it is the only theory that can explain the data available, but rather because it poses specific questions that can be empirically investigated. Efforts in this direction promise to generate ever more detailed and comprehensive accounts of the fundamental components of the visual recognition system and of other cognitive functions.

5.6 Conclusion

Psychologists have been particularly interested in faces because of their high degree of social significance and the special problem that they represent for the visual recognition system. It has been suggested that face processing is subserved by a cognitive module of the type discussed by Fodor (1983). Efforts to support this hypothesis have primarily focused on domain specificity, innateness, and fixed neural architecture. However, other defining properties that are consistent with modularity have received little consideration. The goals of this thesis were to demonstrate that face perception exhibits some of these ancillary characteristics and to use face stimuli to examine aspects of modularity that are less well understood.

The first study indicates that faces automatically trigger holistic processing operations. The composite effect, whereby the bottom half of a face interferes with recognition of the top half of another face, was shown to be applicable to both attended and unattended faces. This study indicates that faces automatically trigger holistic encoding processes.

In a second set of experiments, a novel rivalry phenomenon produced by overlapped upright tilted faces was described. The rivalry effect was investigated on the basis of the participants' ability to clearly perceive and recognize both faces in the stimuli. It was reasoned that perception and recognition of both tilted images in the overlapped displays would be inconsistent with a rivalry effect at short presentation times, but not at longer times where an alternation from one tilted image to the next may occur. The finding that overlapped inverted faces and houses do not produce rivalry was taken as evidence that faces are more readily processed as Gestalts than other complex objects. This interpretation is consistent with the notion that faces engage domain specific operations. These studies also suggest that fast operations underlie perception of a face as a Gestalt. Finally, it was proposed that the rivalry effect produced by overlapped faces may illustrate informational encapsulation in face perception.

In a third set of experiments, faces were used to investigate the relationship between attention and modular functions. Three separate experiments showed that faces and houses compete for attention. This finding suggests that the face perception module does not have its own dedicated resources but rather shares a common pool with other visual processes. Results from one experiment also suggested an advantage for faces in the allocation of attention at very short presentation times. This advantage was interpreted as arising from two interacting mechanisms —faces capturing attention over other objects, and faces being more automatically encoded than other objects.

The thesis strengthens the case for a modular conceptualization of face perception by showing that this function displays several modular characteristics. It also broadens our understanding of the modular architecture proposed by Fodor (1983) by suggesting possible ways that central resources might influence modular processes. A possible avenue for future research in the field of face perception includes the development of more precise definitions for the terms holistic and configural. These definitions should produce consistent results that can be replicated across a wide range of stimulus types and experimental manipulations. It would also be interesting to directly investigate informational encapsulation in face perception and to examine the relationship between this property and some of the face illusions reported in the literature. Finally, whether incorporating attentional mechanisms in face recognition models provides a more complete account of this cognitive function warrants further exploration.

References

Aguirre, G. K., Zarahn, E., & D'Esposito, M. (1998). An area within human ventral cortex sensitive to "building" stimuli: Evidence and implications. *Neuron, 21*, 373-383.

Alais, D., & Blake, R. (1999). Grouping visual features during binocular rivalry. *Vision Research, 39*, 4341-4353.

Andrews, J. T., & Purves, D. (1997). Similarities in normal and binocular rivalrous viewing. *Proceedings of the National Academy of Science, 94*, 9905-9908.

Anstis, S. M. (1975). What does visual perception tell us about visual coding? In M. S. Gazzaniga & C. Blakemore (Eds.), *Handbook of Psychobiology*. New York: Academic Press.

Attneave, F. (1971). Multistability in perception. *Scientific American, 225*, 63-71.

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

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312.

Benso, T. M., Facoetti, A., Galfano, G., Mascetti, G. G., & Umilta, C. (2000). Automatic and voluntary focusing of attention. *Percption & Psychophysics, 62*(5), 935-952.

Biederman, I. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*(1), 38-64.

Biederman, I., & Kalocsai, P. (1998). Neural and psychophysical analysis of object and face recognition. In H. Wechsler & J. P. Phillips & V. Bruce & e. al. (Eds.), *Face recognition: From theory to applications*. Berlin: Springer-Verlag.

Blake, R. (1989). A neural theory of binocular rivalry. *Psychological Review, 96*(1), 145-167.

Bonneh, Y., Sagi, D., & Karni, A. (2001). A transition between eye and object rivalry determined by stimulus coherence. *Vision Research, 41*, 981-989.

Boutet, I., & Chaudhuri, A. (2001). Multistability of overlapped face stimuli is dependent upon orientation. *Perception, 30*(6), 743-753.

Boutet, I., Gentes-Hawn, A., & Chaudhuri, A. (2000). Holistic encoding of composite faces interferes with featural encoding irrespective of attention. *Investigative Ophthalmology & Visual Sciences, 41*(4), S42.

Bradley, A., & Schor, C. (1988). The role of eye movements and masking in monocular rivalry. *Vision Research, 28*(10), 1129-1137.

Bradley, B., Mogg, K., Millar, N., Bonham-Carter, C., Fergusson, E., Jenkins, J., & Parr, M. (1997). Attentional biases for emotional faces. *Cognition and Emotion, 11*, 25-42.

Bradshaw, J. L., & Wallace, G. (1971). Models for the processing and identification of faces. *Perception and Psychophysics, 9*, 443-448.

Braitman, D. (1984). Activity of neurons in monkey posterior temporal cortex during multidimensional visual discrimination task. *Brain Research, 307*, 17-28.

Bruce, V., & Humphreys, G. W. (1994). Recognizing objects and faces. In V. Bruce & G. W. Humphreys (Eds.), *Object and face recognition* (pp. 141-180). Hove, England: Lawrence Erlbaum Associates.

Bruce, V., & Langton, S. (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception, 23*, 803-822.

Bruce, V., Valentine, T., & Baddeley, A. (1987). The basis of the 3/4 view advantage in face recognition. *Applied Cognitive Psychology, 1*, 109-120.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*, 305-327.

Bruyer, R., Laterre, C., Seron, X., Feyereisen, P., Strypstein, E., Pierrard, E., & Recterm, D. (1983). A case of prosopagnosia with some preserved covert remembrance of familar faces. *Brain and Cognition, 2*, 257-284.

Bushnell, I. W. (1982). Discrimination of faces by young infants. *Journal of Experimental Child Psychology, 33*(2), 298-308.

Campbell, F. W., & Howell, E. R. (1972). Monocular alternation: A method for the investigation of pattern vision. *Journal of Physiology, 210*, 19-21.

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

Carey, S., & Diamond, R. (1994). Are faces perceived as configurations more by adults than by children? *Visual Cognition, 1*, 253-274.

Chao, L. L., Martin, A., & Haxby, J. V. (1999). Are face-responsive regions selective only for faces? *Neuroreport, 10*(14), 2945-2950.

Chaudhuri, A. (1990). Modulation of the motion aftereffect by selective attention. *Nature, 344*, 60-62.

Chomsky, N. (1972). *Language and Mind.* New York: Harcourt Brace Jovanovich.

Clark, V. P., Maisog, J. M., & Haxby, J. V. (1998). fMRI study of face perception and memory using random stimulus sequences. *Journal of Neurophysiology, 79*(6), 3257-3265.

Clark, V. P., Parasuraman, R., Keil, K., Kulansky, R., Fannon, S., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1997). Selective attention to face identitiy and color studied with fMRI. *Human Brain Mapping, 5*, 293-297.

Colby, C. L. (1991). The neuroanatomy and neurophysiology of attention. *Journal of Child Neurology, 6*, S90-S118.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience, 22*, 319-349.

Coltheart, M. (1999). Modularity and cognition. *Trends in Cognitive Sciences, 3*(3), 115-120.

Corbetta, M., Miezin, F.M., Dobmeyer, S., Shulman, G.L., & Petersen, S.E. (1991). Selective and divided attention during visual discriminations of shape, color and speed: Functional anatomy by positron emission tomography. *Journal of Neuroscience, 11*, 2383-2402.

Corbetta, M., Miezin, F. M., Dobmeyer, S., Gordon, L., Shulman, S., & Petersen, E. (1991). Attentional modulation of neural processing of shape, color and velocity in humans. *Science, 248*, 1556-1559.

Coull, J. T., Frith, C. D., Buchel, C., & Nobre, A. C. (2000). Orienting attention in time: Behavioural and neuroanatomical distinction between exogenous and endogenous shifts. *Neuropsychologia, 38*(6), 808-819.

Crassini, B., & Broerse, J. (1982). Monocular rivalry occurs without eye movements. *Vision Research, 22*, 203-204.

Creelman, D. C., & Macmillan, N. A. (1991). *Detection theory: A user's guide*. New York: Cambridge Univerisity Press.

D'Esposito, M., Zarahan, E., & Aguirre, G. K. (1999). Event-related functional MRI: Implications for cognitive psychology. *Psychological Bulletin, 125*, 155-164.

Damasio, A. R., Damasio, H. & Van Hoesen, G.W. (1982). Prosopagnosia : Anatomic basis and behavioral mechnisms. *Neurology, 32*, 321-341.

de Gelder, B., & Rouw, R. (2000). Structural encoding precludes recognition of face parts in prosopagnosia. *Cognitive Neuropsychology, 17*, 89-102.

De Haan, E. H. F. (2001). Face perception and recognition, *The handbook of cogntive neuropsychology: What deficits reveal about the human mind* (pp. 75-99). Philadelphia USA: Psychology Press/Taylor & Francis.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*, 193-222.

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

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*(4), 501-517.

Eimer, M. (2000). Attentional modulations of event-related brain potentials sensitive to faces. *Cognitive Neuropsychology, 17*, 103-116.

Ellis, H. D. (1975). Recognizing faces. *British Journal of Psychology, 66*, 409-426.

Ellis, H. D. (1992). The development of face processing skills. In V. Bruce & A. Cowey & A. W. Ellis & D. I. Perrett (Eds.), *Processing of the facial image* (pp. 105-111). Oxford: Clarendon Press.

Engel, E. (1956). The rôle of content in binocular resolution. *American Journal of Psychology, 69,* 87-91.

Enns, J. T., & Shore, D. I. (1997). Separate influences of orientation and lighting direction in the inverted-face effect. *Perception & Psychophysics, 59,* 23-31.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392,* 598-601.

Fahle, M., & Palme, G. (1991). Perceptual rivalry between illusory and real contours. *Biological Cybernetics, 66,* 1-8.

Farah, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigations of domain specificity in visual object recognition. *Current Directions in Psychological Science, 1*(5), 164-169.

Farah, M. J., Rabinowitz, C., Quinn, G. E., & Liu, G. T. (2000). Early commitment of neural substrates for face recognition. *Cognitive Neuropsychology, 17,* 117-123.

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

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. W. (1995b). The inverted face inversion effect in prosopagnosia: Evidence for mandatory face-specific perceptual mechanisms. *Vision Research, 35,* 2089-2093.

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

Fodor, J. (1983). *The Modularity of Mind.* Cambridge, MA: MIT Press.

Fodor, J. (1998). *In Critical Condition.* Cambridge, Massachusets: The MIT Press.

Fodor, J. (2000). *The mind doesn't work that way.* Cambridge, Massachusets: The MIT Press.

Fox, E. (1995). Negative Priming from ignored distractors in visual selection: A review. *Psychonomic Bulletin & Review, 2*(2), 145-173.

Fox, E., Lester, V., Russo, R., Bowles, R. J., Pichler, A., & Dutton, D. (2000). Facial expressions of emotion: Are angry faces detected more efficiently. *Cognition & Emotion, 14*(1), 61-92.

Galper, R. E. (1970). Recognition of faces in photographic negative. *Psychonomic Science, 19*, 207-208.

Galton, F. (1883). *Inquiries into human faculty and its development.* London: MacMillan and co.

Garcia-Perez, M. A. (1992). Eye movements and perceptual multistability. In E. Chekaluk & K. Llewellyn (Eds.), *The role of eye movements in perceptual processes* (pp. 73-109). Amsterdam: Elsevier Science.

Gauthier, I., Behrmann, M., & Tarr, M. J. (1999b). Can face recognition really be dissociated from object recognition? *Journal of Cognitive Neuroscience, 11*(4), 349-370.

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

Gauthier, I., & Tarr, M. J. (1997). Becoming a "greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999a). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*(6), 568-573.

Gauthier, I., Williams, P., Tarr, M. J., & Tanaka, J. (1998). Training "greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38*(15-16), 2401-2428.

Georgeson, M. (1984). Eye movements, afterimages and monocular rivalry. *Vision Research, 24*(10), 1311-1319.

Georgeson, M., & Phillips, R. (1980). Angular selectivity of monocular rivalry: experiment and computer simulation. *Vision Research, 20*, 1007-1013.

Golby, A. J., Gabrieli, J. D., Chiao, J. Y., & Eberhardt, J. L. (2001). Differential responses in the fusiform region to same-race and other- race faces. *Nature Neuroscience, 4*(8), 845-850.

Goren, C. C., Sarty, M., & Wu, P. Y. K. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*, 544-549.

Gregory, R. L. (1970). *The Intelligent Eye.* New York: McGraw-Hill.

Gross, C. G. (1992). Representation of visual stimuli in inferior temporal cortex. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 335*, 3-10.

Gross, C. G., Rocha-Miranda, C. W., & Bender, D. B. (1972). Visual properties of neurons in the inferotemporal cortex of the monkey. *Journal of Neurophysiology, 35*, 96-111.

Grossberg, S. (2000). The complementary brain: unifying brain dynamics and modularity. *Trends in Cognitive Sciences, 4*(6), 233-246.

Hastorf, A. H., Osgood, C. E., & Ono, H. (1966). the semantics of facial expressions and the prediction of the meanings of stereoscopically fused facial expressions. *Scandinavian Journal of Psychology, 7*, 179-188.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science, 293*(5539), 2425-2430.

Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences, 4*(6), 223-233.

Haxby, J. V., Horwitz, B., Ungerleider, L. G., Maisog, J., Pietrini, P., & Grady, C. L. (1994). The functional organization of human extrastriate cortex: A PET-rCBF study of selective attention to faces and locations. *Journal of Neuroscience, 14*, 6336-6353.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron, 22*(1), 189-199.

Hayes, A. (1988). Identification of two-tone images; some implications for high- and low-spatial frequency processes in human vision. *Perception, 17*(4), 429-436.

Hayes, T., Morrone, M. C., & Burr, D. C. (1986). Recognition of positive and negative bandpass-filtered images. *Perception, 15*, 595-602.

Heywood, C. A., & Cowey, A. (1992). The role of the 'face-cell' area in the discrimination and recognition of faces by monkeys. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 335*, 31-37.

Hill, H., & Bruce, V. (1993). The effects of lighting on the perception of facial surfaces. *Journal of Experimental Psychology: Human Perception & Performance, 22*, 986-1004.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23*, 65-74.

Hole, G. J., George, P. A., & Dunsmore, V. (1999). Evidence for holistic encoding of faces viewed as photographic negatives. *Perception, 28*, 341-359.

Homa, D., Haver, B., & Schwartz, T. (1976). Perceptibility of schematic face stimuli: Evidence for a perceptual Gestalt. *Memory & Cognition, 4*(2), 176-185.

Houghton, G., Tipper, S.P., Weaver, B., & Shore, D.I. (1996). Inhibition and interference in selective attention: Some tests of a neural network model. *Visual Cognition, 3*, 11-164.

Ishai, A., Ungerleider, L., Martin, A., Maisog, J. M., & Haxby, J. V. (1997). fMRI reveals differential activation in the ventral object vision pathway during the perception of faces, houses and chairs. *Neuroimage, 5*, S149.

Ishai, A., Ungerleider, L. G., Martin, A., Schouten, J. L., & Haxby, J. V. (1999). Distributed representation of objects in the human ventral visual pathway. *Proceedings of the National Academy of Science USA, 96*(16), 9379-9384.

Jeffreys, D. A. (1996). Evoked potential studies of face and object processing. *Visual Cognition, 3*, 1-38.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborn's preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40*, 1-19.

Johnston, A., Hill, H., & Carman, N. (1992). Recognising faces: Effects of lighting direction. *Perception, 21*(3), 1992.

Kanwisher, N. (2000). Domain specificity in face perception. *Nature Neuroscience, 3*(8), 759-763.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neurscience, 17*(11), 4302-4311.

Kanwisher, N., Stanley, D., & Harris, A. (1999). The fusiform face area is selective for faces not animals. *Neuroreport, 10*(1), 183-187.

Kanwisher, N., Tong, F., & Nakayama, K. (1998). The effect of face inversion on the human fusiform face area. *Cognition, 68*(1), B1-11.

Kemp, R., McManus, C., & Pigott, T. (1990). Sensitivity to the displacement of facial features in negative and inverted images. *Perception, 19*(4), 531-543.

Khurana, B. (2000). Not to be and then to be: Visual representation of ignored unfamilar faces. *Journal of Experimental Psychology: Human Perception & Performance, 26*(1), 246-263.

Kitterle, F. L., & Thomas, J. (1980). The effects of spatial frequency, orientation and color upon binocular rivalry and monocular pattern alternation. *Bulletin of the Psychonomic Society, 16*, 405-407.

Kobatake, E., Wang, G., & Tanaka, K. (1998). Effects of shape-discrimination training on the selectivity of inferotemporal cells in adult monkeys. *Journal of Neurophysiology, 80*, 324-330.

Kovacs, I., Papathomas, T. V., Yang, M., & Féher, A. (1996). When the brain changes its mind: Interocular grouping during binocular rivalry. *Proceedings of the National Academy of Science, 93*, 15508-15511.

Kristofferson, M. (1972). When item recognition and visual search functions are similar. *Perception & Psychophysics, 12*.

Kuehn, S. M. (1994). Impact of quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception, 23*, 95-122.

Langton, S. R. H., & Bruce, V. (1999). Reflexive visual orienting in response to the social attention of others. *Visual Cognition, 6*, 541-568.

Lavie, N., & Fox, E. (2000). The role of perceptual load in negative priming. *Journal of Experimental Psychology: Human Perception and Performance, 26*(3), 1038-1052.

Leder, H. (1996). Line drawings of faces reduce configural processing. *Perception, 25*, 355-366.

Leder, H., & Bruce, V. (2000). When inverted faces are recognized: The role of configural information in face recognition. *The Quarterly Journal of Experimental Psychology, 53A*(2), 513-536.

LeGrand, R., Mondloch, C. J., Maurer, D., & Brent, H. P. (2001). Early visual experience and face processing. *Nature, 410*, 890.

Leopold, D. A., & Logothetis, N. K. (1996). Activity changes in early visual cortex reflect monkey's percepts during binocular rivalry. *Nature, 379*, 549-553.

Lewis, M. B., & Johnston, R. A. (1997). The Thatcher illusion as a test of configural disruption. *Perception, 26*, 225-227.

Liu, C. H., & Chaudhuri, A. (1997). Face recognition with multi-ton and two-tone photographic negatives. *Perception, 26*, 1289-1296.

Liu, C. H., Collin, C. A., Rainville, S. J., & Chaudhuri, A. (2000). The effects of spatial frequency overlap on face recognition. *Journal of Experimental Psychology: Human Perception and Performance, 26*(3), 956-979.

Logothetis, N. K. (2000). Object recognition: Holistic representations in the monkey brain. *Spatial Vision, 13*, 165-178.

Long, G. M., & Olszweski, A. D. (1999). To reverse or not to reverse: When is an ambiguous figure not ambiguous? *American Journal of Psychology, 112*(1), 41-71.

Long, G. M., & Toppino, T. C. (1981). Multiple representations of the same reversible figure: implications for cognitive descisional interpretatons. *Perception & Psychophysics, 10*, 231-234.

Lumer, E. D., Firston, K. J., & Rees, G. (1998). Neural correlates of perceptual rivalry in the human brain. *Science, 280*, 1930-1934.

Mack, A., & Rock, I. (1998). *Inattentional Blindness*. Cambridge MA: MIT Press.

Malpass, R. S. (1981). Training in face recognition. In G. Davies & H. D. Ellis & J. Shepherd (Eds.), *Perceiving and remembering faces* (pp. 271-285). New York: Academic Press.

Mapperson, B., & Lovegrove, W. (1978). Spatial frequency and orientation specificity of simultaneous and successive inhibition on monocular rivalry. *Vision Research, 18,* 1461-1463.

Mapperson, B., & Lovegrove, W. (1984). The dependence of monocular rivalry on spatial frequency: Some interaction variables. *Perception, 13,* 141-152.

May, C. P., Kane, M.J., & Hasher, L. (1995). Determinants of negative priming. *Psychological Bulletin, 118,* 35-54.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9*(5), 605-610.

McNeil, J. E., & Warrington, E. K. (1993). Prosopagnosia: A face-specific disorder. *The Quarterly Journal of Experimental Psychology, 46A*(1), 1-10.

Mondor, T. A., & Amirault, K. J. (1998). Effect of same- and different-modality spatial cues on auditory and visual target identification. *Journal of Experimental Psychology: Human Perception & Performance, 24*(3), 745-755.

Moran, J., & Desimone, R. (1985). Selective Attention Gates Visual Processing in the Extrastriate Cortex. *Science, 229,* 782-784.

Morton, J., & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review, 98*(2), 164-181.

Moscovitch, M., & Moscovitch, D. A. (2000). Super face-inversion effects for isolated internal or external features, and for fractured faces. *Cognitive Neuropsychology, 17,* 201-219.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? Nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Journal of Cognitive Neuroscience, 9*(5), 555-604.

Motter, B. C. (1994). Neural correlates of attentive selection for color or luminance in extrastriate area V4. *The Journal of Neuroscience, 14*(4), 2178-2189.

Motter, C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2 and V4 in the presence of competing stimuli. *Journal of neurophysiology, 30*(3), 909-919.

Mukai, I., & Watanabe, T. (2001). Differential effect of attention to translation and expansion on motion aftereffects (MAE). *Vision Research, 41*(9), 1107-1117.

Neisser, U. (1967). *Cognitive psychology.* New York: Appleton-Century-Crofts.

Neisser, U., & Becklen, R. (1975). Selective looking: Attending to visually specified events. *Cognitive Psychology, 7,* 480-494.

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology, 7,* 44-64.

Nothdurft, H. C. (1993). Faces and facial expressions do not pop out. *Perception, 22,* 1287-1298.

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature, 401,* 584-587.

Öhman, A., Lundqvist, D., & Esteves, F. (2001). The face in the crowd revisited: A threat advantage with schematic stimuli. *Journal of Personality & Social Psychology, 80*(3), 381-396.

Ono, H., Hastorf, A. H., & Osgood, C. E. (1966). Binocular rivalry as a function of incongruity in meaning. *Scandinavian Journal of Psychology, 7,* 225-233.

Palermo, R., & Rhodes, G. (2001). The influence of divided attention on holistic face perception. *Cognition, In Press.*

Pelton, L. H., Solley, C. M., & Brent, S. B. (1969). Stability and attraction of perceptual organizations in the reversible figure-ground phenomenon. *Perceptual & Motor Skills, 1969*(28), 983-991.

Penhune, V. B., Zatorre, R. J., & Evans, A. C. (1998). cerebellar contributions to motor timing: A PET study of auditory and visual rhytm reproduction. *Journal of Cognitive Neuroscience, 10,* 752-766.

Perrett, D. I., Hietanen, J. K., Oram, M. W., & Benson, P. J. (1992). Organization and functions of cells responsive to faces in the temporal cortex. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 335*, 23-30.

Perrett, D. I., Mistlin, A. J., & Chitty, A. J. (1987). Visual neurones responsive to faces. *Trends in Neuroscience, 10*(9), 358-364.

Phillips, R. J. (1972). Why are faces hard to recognize in photographic negative? *Perception & Psychophysics, 12*, 425-426.

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual Review of Neuroscience, 13*, 25-42.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: a functional magnetic resonance imaging study. *Journal of Neuroscience, 16*(16), 5205-5215.

Puce, A., Truett, A., Gore, J. C., & McCarthy, G. (1995). Face-sensitive regions in human extrastriate cortex studied by functional MRI. *Journal of Neurophysiology, 74*(3), 1192-1199.

Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: Configuration enhances detection. *Perception & Psychophysics, 43*, 355-366.

Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception, 25*(9), 1091-1108.

Rauschecker, J. P. J., Campbell, F. W., & Atkinson, J. (1973). Colour opponent neurones in the human visual system. *Nature, 245*, 42-43.

Reinitz, M. T., Bartlett, J. C., & Searcy, J. (1997). Role of attention in encoding facial features and their spatial relations. *Paper presented at the Psychonomic Society Conference, Philadelphia, Pennsylvania.*

Reinitz, M. T., Morrissey, J., & Demb, J. (1994). Role of attention in face encoding. *Journal of experimental psychology: Learning, memory and cognition, 20*(1), 161-168.

Reisberg, D., & O'Shaughnessy, M. (1984). Diverting subjects' concentration slows figural reversal. *Perception & Psychophysics, 13*, 461-468.

Rhodes, G. (1988). Looking at faces: First-order and second-order features as determinants of facial appearance. *Perception, 17,* 43-63.

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

Rock, I. (1975). *An Introduction to Perception.* New York: MacMillan.

Rock, I. (1985). Peception and knowledge. *Acta Psychologica, 59,* 3-22.

Rock, I., & Mitchener, K. (1992). Further evidence of failure of reversal of ambiguous figures by uninformed subjects. *Perception & Psychophysics, 21,* 39-45.

Rolls, E. T. (1992). Neurophysiological mechanisms underlying face processing within and beyond temporal cortical visual areas. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 335,* 11-21.

Ruggieri, B., & Fernandez, M. F. (1994). Gaze orientation in perception of reversible figures. *Perceptual & Motor Skills, 78,* 299-303.

Saumier, D., Arguin, M., & Lassonde, M. (2001). Prosopagnosia: A case study involving problems in processing configural information. *Brain and Cognition, 46,* 255-316.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: 1. Detection, search, and attention. *Psychological Review, 84.*

Schyns, P. G., & Bulthoff, H. H. (1994). Viewpoint dependence and face recognition. *Proceedings of the XVI Meeting of the Cognitive Science Society,* 789-793.

Scotto, M. A., Oliva, G. A., & Tuccio, M. T. (1990). Eye movements and reversal rates of ambiguous patterns. *Perceptual & Motor Skills, 70,* 1059-1073.

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

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75,* 221-242.

Sergent, J., & Signoret, J. L. (1992). Varieties of functional deficits in prosopagnosia. *Cerebral Cortex, 2*(5), 375-388.

Sheinberg, D. L., & Logothetis, N. K. (1997). The role of temporal cortical areas in perceptual organization. *Proceedings of the National Academy of Science, 94*, 3408-3413.

Shiffrin, R. M. (1988). Attention. In R. C. Atkinson & R. J. Herrinstein & G. Lindzey & R. D. Luce (Eds.), *Steven's handbook of experimental psychology* (Vol. 2). New York: Wiley.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: 2. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*, 127-190.

Stankiewicz, B. J., Hummel, J.E., and Cooper, E.E. (1998). The role of attention in priming for left-right reflections of object images: Evidence for a dual representation of object shape. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 732-744.

Subramaniam, & Biederman, I. (1997). Does contrast reversal affect object identification. *Investigative Ophthalmology & Visual Science, 38*, 998.

Sugiura, M., Kawashima, R., Nakamura, K., Sato, N., Nakamura, A., Kato, T., Hatano, K., Schormann, T., Zilles, K., Sato, K., Ito, K., & Fukuda, H. (2001). Activation reduction in anterior temporal cortices during repeated recognition of faces of personal acquaintances. *Neuroimage, 13*(5), 877-890.

Suzuki, S., & Cavanagh, P. (1995). Facial organization blocks access to low-level features: An object inferiority effect. *Journal of Experimental Psychology: Human Perception & Performance, 21*(4), 901-913.

Tanaka, J., & Gauthier, I. (1997). Expertise in object and face recognition. *The Psychology of Learning and Motivation, 36*, 83-125.

Tanaka, J. N., Kay, J. B., Grinnell, E., Stansfield, B., & Szechter, L. (1998). Face recognition in young children: When the whole is greater than the sum of its parts. *Visual Cognition, 5*, 479-496.

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46A*(2), 225-245.

Tanaka, J. W., & Sengco, J. A. (1997). Features and their configuration in face recognition. *Memory & Cognition, 25*, 583-592.

Teuber, H. L. (1968). Disorders of memory following penetrating missile wounds of the brain. *Neurology, 18*, 287-288.

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology, 37A*, 571-590.

Tipper, S. P. (2001). Does negative priming reflect inhibitory mechanisms? A reviiew and integration of conflicting views. *The Quarterly Journal of Experimental Psychology, 54A*(2), 321-343.

Tipper, S. P., & Driver, J. (1988). Negative priming between pictures and words in a selective attention task: Evidence for semantic processing of ignored stimuli. *Memory & Cognition, 16*(1), 64-70.

Tippett, L. J., Miller, L. A., & Farah, M. J. (2000). Prosopamnesia: A selective impairment in face learning. *Cognitive Neuropsychology, 17,* 241-255.

Tong, F., & Engel, S. A. (2001). Interocular rivalry revealed in the human cortical blind-spot representation. *Nature, 411,* 195-1999.

Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of Experimental Psychology: Human Perception & Performance, 25*(4), 1016-1035.

Tong, F., Nakayama, K., Moscovitch, M., Weinrib, O., & Kanwisher, N. (2000). Response properties of the human fusiform face area. *Cognitive Neuropsychology, 17,* 257-279.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21,* 753-759.

Toppino, T. C., & Long, G. M. (1987). Selective adaptation with reversible figures: Don't change that channel. *Perception & Psychophysics, 42*(1), 37-48.

Tovée, M. J. (1998). Is face processing special? *Neuron, 21,* 1239-1242.

Treisman, A. (1988). Features and objects: The fourteenth Bartlett Memorial Lecture. *Quarterly Journal of Experimental Psychology, 40A,* 201-237.

Treisman, A., & DeSchepper, B. (1996). Object tokens, attention, and visual memory. In T. Inui (Ed.), *Attention and performance 16: Information integration in perception and communication* (pp. 15-46). Cambridge: Mit Press.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception & Performance, 16*(3), 459-478.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology, 12*, 97-136.

Troje, N. F., & Bülthoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Visual Research, 36*(12), 1761-1771.

Tsal, Y., & Kolbet, L. (1985). Disambiguating ambiguous figures by selective attention. *The Quarterly Journal of Experimental Psychology, 37A*, 25-37.

Tsunoda, K., Yamane, Y., Nishizaki, M., & Tanifuji, M. (2001). Complex objects are represented in macaque inferotemporal cortex by the combination of feature columns. *Nature Neuroscience, 4*(8), 832-838.

Turatto, M., Benso, F., Facoetti, A., Galfano, G., Mascetti, G. G., & Umilta, C. (2000). Automatic and voluntary focusing of attention. *Perception & Psychophysics, 62*(5), 935-952.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79*, 471-491.

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *Quarterly Journal of Experimental Psychology A, 43*(2), 161-204.

Vuilleumier, P. (2000). Faces call for attention: evidence from patients with visual extinction. *Neuropsychologia, 38*, 693-700.

Vuilleumier, P., Armony, J. L., Driver, J., & Dolan, R. J. (2001). Effect of attention and emotion on face processing in the human brain: An event-related fMRI study. *Neuron, 30*, 829-841.

Vuilleumier, P., & Schwartz, S. (2001). Emotional facial expressions capture attention. *Neurology, 56*, 153-158.

Walker, P. (1976). The perceptual fragmentation of unstablized images. *Quarterly Journal of Experimental Psychology, 28*, 35-45.

Wallis, G., & Rolls, E. T. (1997). Invariant face and object recognition in the visual system. *Progress in Neurobiology, 51*, 167-194.

Walton, G. E., Bower, N. J. A., & Bower, T. G. R. (1992). Recognition of familiar faces by newborns. *Infant Behavior and Development, 15*, 265-269.

Whalen, P. J., Rauch, S. L., Etcoff, N. L., McInerney, S. C., Lee, M. B., & Jenike, M. A. (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. *Journal of Neuroscience, 1*(1), 411-418.

White, M. (1995). Preattentive analysis of facial expressions and emotions. *Cognition and Emotion, 9*, 439-460.

Wojciulik, E., Kanwisher, N., & Driver, J. (1998). Covert visual attention modulates face-specific activity in the human fusiform gyrus: fMRI study. *Journal of Neurophysiology, 79*(3), 1574-1578.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81*, 141-145.

Young, A. W. (1992). Face recognition impairments. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 335*, 47-54.

Young, A. W., Hellawell, D., & Hay, D. C. (1987). Configural information in face perception. *Perception, 16*, 747-759.

Young, A. W., Newcombe, F., de Haan, E. H. F., Small, M., & Hay, D. C. (1993). Face perception after brain injury: selective impairments affecting identity and expression. *Brain, 116*, 941-959.

Yu, K., & Blake, R. (1992). Do recognizable figures enjoy and advantage in binocular rivalry? *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1158-1173.

Zeki, S., & Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions of the Royal Society of London B Biological Sciences, 353*(1377), 1911-1914.