UNIVERSITY OF SOUTHAMPTON

FACULTY OF SOCIAL AND HUMAN SCIENCES

Psychology

Identifying Sources of Configural Face Processing

by

Natalie Mestry

Thesis for the degree of Doctor of Philosophy

December 2012

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF SOCIAL AND HUMAN SCIENCES

Psychology

Thesis for the degree of Doctor of Philosophy

IDENTIFYING SOURCES OF CONFIGURAL FACE PROCESSING

Natalie Mestry

Absence of a precise definition of configural processing in face perception has resulted in previous demonstrations now being accounted for by decisional rather than perceptual processes (Wenger & Ingvalson, 2002, 2003; Richler, Gauthier, Wenger, & Palmeri, 2008; Cornes, Donnelly, Godwin, & Wenger, 2011). In this thesis, I show that many of the difficulties in discriminating between competing accounts of configural processing result from an incomplete mapping between theoretical frameworks, experiments and data. Furthermore, by using general recognition theory (GRT, Ashby & Townsend, 1986) to make the mapping more complete, I demonstrate perceptual and decisional sources of configurality across three face processing tasks. GRT provides formal definitions for the ways in which multiple stimulus dimensions can interact, and thus provides a framework for modelling the dependencies that are indicative of configural processing. Changes to feature size, feature identity and feature orientation have been explored within the GRT framework. For one of these tasks, The Feature Orientation Task which is analogous to the Thatcher illusion (Thompson, 1980), evidence for three types of dependencies exists. Converging evidence for multiple sources of configurality was provided when stimuli from this Thatcher illusion task were used in an event-related potential (ERP) study. The ERP task revealed evidence for a decisional effect and a mapping between the GRT violations and the ERP effects identified across the face components. Finally, the value in applying GRT to specific populations is demonstrated in studies of development and prosopagnosia. Overall, this thesis demonstrates there are multiple sources of configural face processing and the GRT paradigm is helpful in understanding the development, stability and impairments of these sources.

Contents

List of Tables

List of Figures

Declaration of Authorship

I, *Natalie Mestry* declare that the thesis entitled *Identifying Sources of Configural Face Processing* and the work presented in it are my own and has been generated by me as the result of my own original research. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University;
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- Either none of this work has been published before submission, or parts of this work have been published as:

Mestry, N., Menneer, T., Wenger, M. J., & Donnelly, N. (2012). Identifying sources of configurality in three face processing tasks. *Frontiers in Psychology, 3,* 456.

Mestry, N., Donnelly, N., Menneer, T., & McCarthy, R. A. (2012). Discriminating Thatcherised from typical faces in a case of prosopagnosia. *Neuropsychologia, 50,* 3410-3418.

Signed: ………………………………………………………………………..

Date:……………………………………………………………………………..

Acknowledgements

There are many people who have helped along the way and the work in this thesis would not have been possible without their support. Firstly, I would like to thank my supervisor, Nick Donnelly, for his guidance, ideas and the opportunity to work on such an interesting project. I am also very grateful to Tamaryn Menneer for her help throughout my PhD. The support of these two individuals was invaluable and is something for which I am forever indebted.

I would like to give special mention to those who have offered their expertise with particular projects: Michael Wenger for his insightful comments and enthusiasm in discussions about GRT; Nick Benikos, Samantha Broyd and Ben Ainsworth for their time and guidance with the ERP technique; Erich Graf for help creating the ERP stimuli and Rosaleen McCarthy and those who have supported the prosopagnosia project.

I would like to thank the CVC group and all my office mates over the years for their time and advice, especially Hayward Godwin and Katie Gray for their wise words and programming expertise in the early stages of my PhD. Special thanks to the lunch-break group who brightened the darker days and to my taekwondo club who have provided the much needed escapism but also the focus to continue.

I am also very grateful to those who took part in my (often very long) experiments! Finally, I would like to thank my family for their support and especially David for helping me get to the end with constant support, advice and love.

The research in this thesis was funded by the Economic and Social Research Council.

Abbreviations

2AFC	two-alternative force choice
c	criterion (bias)
CPA	complete over part probe advantage
d'	*d*-prime (sensitivity)
DS	decisional separability
EEG	electrocephalogram
EOG	electrooculogram
ERP	event-related potential
FA	false alarm
fMRI	functional magnetic resonance imaging
GFP	global field power
GRT	general recognition theory
HR	hit rate
MRI	magnetic resonance imaging
ms	milliseconds
MSDA	multidimensional signal detection analysis
PHD	ID for participant with apperceptive prosopagnosia
PI	perceptual independence
PS	perceptual separability
RGB	red, green, blue
RI	retention interval
RMS	root mean square
RT	reaction time
SDT	signal detection theory

Chapter 1

Introduction

Faces are structurally similar and yet differentiating between most faces is thought to be rather easy. However, the processes that allow us to individuate faces have not been adequately defined. This has led to the development of differing hypotheses regarding how information within a face is represented. For typical adults, upright faces are thought to be processed as wholes (e.g. Farah, Wilson, Drain, & Tanaka, 1998) or configurations (e.g. Bartlett & Searcy, 1993) or both (Maurer, Le Grand, & Mondloch, 2002), as well as by features. For upright faces to be processed as wholes, their representation would involve little, if any, decomposition into parts. This *holistic processing* would then be disrupted when faces are inverted (holistic perception hypothesis, Farah et al., 1998). For upright faces to be processed as configurations, there would be two sources of processing: sensitivity to *configural information* (spatial relations between parts) and featural processing. In this case, when a face is inverted the configural information is disrupted so that processing becomes feature-based (dual-mode hypothesis, Searcy & Bartlett, 1996).

Determining how faces are processed requires experiments capable of differentiating between competing accounts. In this thesis, I examine whether the experiments and data typically thought to support one kind of account versus another are, in fact, subject to multiple interpretations. Given the differences between the familiar and unfamiliar face processing (Burton & Jenkins, 2011), with people better at recognising familiar faces (Hancock, Bruce & Burton, 2000), this issue is examined with reference to the processing of unfamiliar faces only.

Mapping between Theoretical Frameworks, Experiments and Data in Face Processing.

Many studies have explored whether there is some form of cognitive process that is unique to face processing. This question reduces to whether faces are processed or represented differently from all other objects. Previous studies have addressed this question by exploring if processing of faces is especially subject to (1) influence of orientation, (2) a failure of selective attention to features or (3) an enhanced sensitivity to the spatial positioning of features (configural information).

The Effects of Inversion on Face and Object Processing

Yin (1969) examined whether recognition impairment exists for mono-oriented objects when they are inverted. Participants were presented with a series of pictures to remember that included faces, houses and airplanes. In a test phase, participants were presented with a pair of pictures and indicated which of each pair was presented in the encoding phase. Object orientation was varied between encoding and test phases. Yin (1969) reported that faces were disproportionately impaired by inversion compared to houses and airplanes and suggested that there is something special about faces that compromised recognition of inverted faces.

The ubiquitous nature of this finding has been reflected in multiple follow-up studies (e.g. Scapinello & Yarmey, 1970; Yarmey, 1971; Carey & Diamond, 1977; Diamond & Carey, 1986; Valentine & Bruce, 1986). The magnitude of the face inversion effect has become, for many, a diagnostic measure illustrating the unique nature of face processing (e.g. Yin, 1969; Carey & Diamond, 1977; Diamond & Carey, 1986). However, there are two problems with this argument. The first is that recent studies have shown that inversion effects can be found with other categories of objects in which we are experts. For example, dog breeders and judges who had experience with dogs showed the inversion effect for recognising breads of dogs, but novice participants did not (Diamond & Carey, 1986). Experts had higher accuracy than novices when recognising breeds of dogs in their upright orientation, but lower accuracy than novices when recognising breeds of dogs in their inverted orientation. These findings demonstrate that the increased inversion effect found with faces versus objects is removed when expertise and familiarity is matched. Therefore, these data confirm that inversion effects are not face specific.

The second problem is the form of the interaction between orientation and stimulus type. For an interaction to be used to support different representations of objects and faces, it must be a crossover interaction. In the absence of a crossover interaction, the differential effect of orientation can be interpreted as a scaling effect. Unfortunately, the form of interaction across stimulus type and orientation is rarely subject to real scrutiny. Where it has been, the interaction is usually not a crossover interaction (e.g. Rouse, Donnelly, Hadwin, & Brown, 2004). Although stated in a different way, this is essentially the criticism levelled by Valentine (1988) for face inversion effects; the effects of inversion are quantitative, representing an increase in the 'error' of encoding rather than a shift between modes of processing. Therefore, it must be stressed that even though faces are typically more affected by inversion than objects, this is not definitive evidence for faces being processed in a qualitatively different manner to objects.

The Failure of Selective Attention to Parts in Faces

The Part-Whole Effect

Tanaka and Farah (1993) explored the part and whole face recognition in a two-alternative forced choice paradigm for faces and scrambled faces, upright and inverted faces, and faces and houses. Participants found it harder to identify part of a face in isolation compared to in the context of a whole face (the part-whole effect, Figure 1.1). However, there was no disadvantage found for part identification in isolation for inverted faces, scrambled faces or houses. These results were used to suggest holistic representation of upright faces, but part-based representation of inverted faces and objects (in this case houses).

Not all findings using variants of the part-whole task show this level of clarity. Davidoff and Donnelly (1990) compared complete and part probes in a face matching task similar to that of Tanaka and Farah (1993). An advantage was found for complete probes over part probes for face stimuli. These results indicated holistic information was advantageous in the matching process, consistent with Tanaka and Farah (1993). However, this *complete over part probe advantage* (CPA) was also found for chairs (their choice of familiar objects), but not for scrambled faces or scrambled chairs. Tanaka and Farah (1993) found no disadvantage for part identification in isolation for houses. Donnelly and Davidoff (1999) explored why differential CPA results were found for chairs and houses across these two studies, suggesting features are represented holistically for both upright faces and chairs but not houses. Evidence was consistent with the holistic perception hypothesis for faces (Farah et al., 1998; Davidoff & Donnelly, 1990; Donnelly & Davidoff, 1999; Tanaka & Farah, 1993). However, it was also found to be inconsistent with the holistic perception hypothesis; familiar objects can show CPA too and can therefore be represented holistically (Davidoff & Donnelly, 1990; Donnelly & Davidoff, 1999). Thus, evidence for differences between face and object processing from part-whole studies are likely to be based on quantitative rather than qualitative effects.

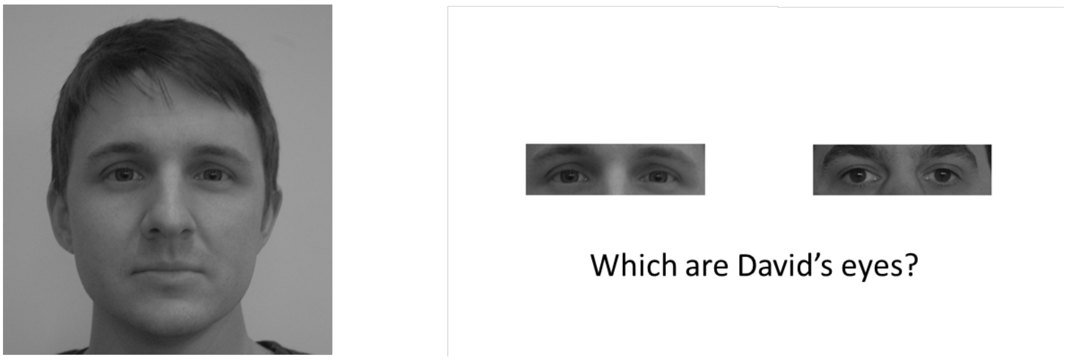


Figure 1.1. Figure demonstrating the type of stimuli used in the part-whole studies. It is easier to identify part of a face in the context of a whole face than in isolation.

Gauthier and Tarr (1997) used a part recognition paradigm to demonstrate that inversion effects are not unique to faces and could be acquired with training. In their task, novice participants were shown artificial patterns called greebles and learnt the names of three parts to these objects. Novice participants were shown six different greebles and then had to identify parts belonging to greebles in a forced choice paradigm. The parts were either shown in the correct context, incorrect context or in isolation. The expert group were given training to become familiar with greebles prior to the task. Familiarity with greebles led to increased accuracy, reaction speed and sensitivity to configural changes in greebles for experts compared to novices. Critically, for the experts, recognition of upright greeble parts was disrupted (longer reaction time, RT) when presented in different configural contexts relative to the correct configural context. When the greebles were inverted, the experts no longer showed this difference. Novices showed no such contrast across orientation. Therefore, in line with evidence in the previous section, the inversion effect in part-whole studies is not face specific.

The Tanaka and Farah (1993) study was extended in a series of related studies. Farah et al. (1998) reported four experiments. In Experiment 1, two faces were displayed that were the same or different in one or more features. Same-different judgments were made based on a specific target feature. Therefore, in this experiment, it was possible that non-target features could also vary. The experiment explored the influence of changes in an irrelevant feature on target feature recognition. An advantage was found for correct responses to the target feature when non-target features were compatible (when both were the same or different) in upright faces. However, this was not the case for inverted faces, scrambled faces or houses. Farah et al. (1998) concluded that this showed processing of upright faces involves relatively little part decomposition.

In a further same-different matching task (Experiment 2), Farah et al. (1998) created a paradigm analogous to that used by Johnston and McClelland (1980) for investigating the effect of hierarchical processing for recognition of words. A mask stimulus was presented between the study and test face. The mask was either a whole face or part of a face (neither was composed of parts used in the stimuli for that trial). Word stimuli were also used as a non-face comparison, with whole and part word masks. Farah et al. (1998) found that upright faces were more disrupted by whole face masks than part face masks (and were also more affected than the word stimuli). Farah et al. (1998) interpreted this finding as the parts of upright faces not being explicitly represented and, therefore, not affected by the intervening face parts mask. Inverted faces (Experiment 3) were used to demonstrate that the differences shown between the face and word stimuli were not due to low level differences, as accuracy for inverted faces failed to show a difference between the mask conditions. Houses (Experiment 4) were used as an intermediate object between faces and words, demonstrating an intermediate effect of mask type. Farah et al. (1998) suggest this series of studies shows evidence for the holistic representation of faces: that faces are processed as wholes and parts are not explicitly represented. The results also suggest that processing faces show a special degree of reliance on holistic representation not a qualitative difference in processing from objects as the discrepancy between the word and house control conditions is interpreted as an indication of the degree of holistic representation.

Tanaka and Sengco (1997) demonstrated evidence for interdependency of featural and configural information in a face. In this study, participants studied upright faces that had spacing manipulated between features and were then tested on recognition for the features either in isolation, in a new configuration (the feature spaced differently in the face context) or in the old face configuration. The results showed recognition was best for features in the old configuration and least successful for testing the part in isolation. For inverted faces and non-face stimuli there was no difference in accuracy between the different test conditions.

Together, these studies examining interdependency between features using the part-whole paradigm suggest processing upright faces involves relatively little part decomposition, in support of the holistic perception hypothesis. The findings from Farah et al. (1998) Experiment 1, which suggested increased accuracy when non-target parts were consistent with the same/different status of the target feature, have been replicated widely. However, later work has shown these effects are underpinned by changes in bias, a decisional component, rather than just a change in perceptual sensitivity[1] (Wenger & Ingvalson, 2002). This is an important distinction as the holistic perception hypothesis is built on an assumption that holism is perceptual in nature.

The Composite Face Effect

Young, Hellawell and Hay (1987) manipulated the arrangement of the top and bottom halves of faces such that they were aligned or misaligned. Composite faces were formed from different faces (Figure 1.2). The question of interest was the effect of alignment on the RT to recognise part faces. The composite face effect refers to the slower and less accurate recognition of face halves in upright aligned relative to misaligned faces (Young et al., 1987). The effect is usually interpreted as showing face halves being bound together holistically when aligned, but not when misaligned. The

[1] Sensitivity and bias as separate perceptual and decisional effects are outlined in more detail in Appendix A in the context of signal detection theory (SDT, Green & Swets, 1966).
[2] See appendix A for distinction of sensitivity and bias as perceptual and decisional effects
[3] Mauchly's sphericity test could not be computed for any of these factors as df = 1 (i.e, if k = 2). When k = 2, sphericity is always met as there is only one variance of the difference between the levels.

irrelevant face half is believed to influence the target half, leading to the difference in performance in the aligned condition. For inverted composite faces, no difference in performance between aligned and misaligned faces was found (Young et al., 1987). This is interpreted as evidence for an absence of holism for inverted faces. The composite face effect has been replicated many times (e.g. Hole, 1994; Richler, Gauthier, Wenger, & Palmeri, 2008; Goffaux & Rossion, 2006) and the effect is robust. It is larger when faces are famous and recognisable, than when the top half is unfamiliar and learnt along with an arbitrary surname (Young et al., 1987).

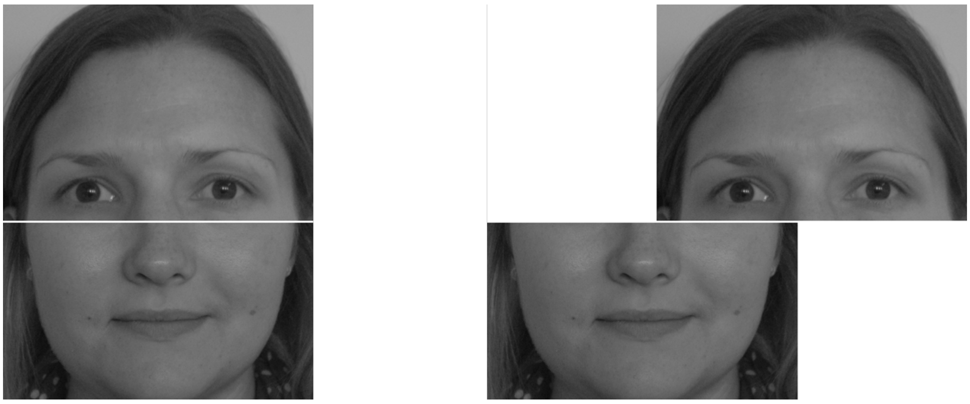


Figure 1.2. Figure demonstrating the type of stimuli used in composite face studies. The top face half is easier to recognise when the parts are misaligned than aligned.

The recognition task conducted by Young et al. (1987) used what has become known as the *partial design*. In the partial design, the irrelevant face half in the test face is always different from the study face, but the influence of this non-target part is not taken into account in analysis. Therefore, it is possible that irrelevant face halves influence decisions and not perceptions. As only accuracy is usually measured, response bias[2] for aligned versus misaligned trials could influence performance (Richler, Mack, Palmeri, & Gauthier, 2011). Consistent with this idea, Richler, Cheung and Gauthier (2011a) demonstrated the partial design is susceptible to response biases that impact on whether evidence for selective attention is observed.

To deal with the issue of response bias in relation to the partial design, a complete design can be formed by including both same and different irrelevant face halves in the test face. Holistic processing is then inferred from the congruency effect

[2] See appendix A for distinction of sensitivity and bias as perceptual and decisional effects

where performance is better when both the target part and non-target part are either the same or different (congruent), than when one is the same and one is different (incongruent). Performance is measured in terms of sensitivity and bias (e.g. Richler et al., 2008; Cheung & Gauthier, 2010; Richler, Mack et al., 2011; Richler, Cheung, et al., 2011a). Using the complete design, evidence points to decisional processes underlying the composite face effect (Richler et al., 2008).

Sensitivity to Spatial Positioning of Features

The Thatcher Illusion

The *Thatcher illusion* (Thompson, 1980) is formed when the eyes and mouth in a face are 'Thatcherised' (inverted relative to the context) leading the face to appear grotesque. When the Thatcherised face is then inverted, the grotesque appearance disappears (Figure 1.3). The change in appearance was shown to occur quite abruptly between 94-100 degrees of rotation in a study examining angle at which the perception changed (Stürzel & Spillmann, 2000) and when orientated stimuli were rated for grotesqueness (Murray, Yong, & Rhodes, 2000). However, in a reaction time study to identify rotated faces as either Thatcherised or normal, evidence for a gradual change in processing was found (Lewis, 2001).

The grotesque appearance of an upright Thatcherised face has been attributed to the perception of manipulated configural information in upright faces. The grotesque appearance is reduced when configural information is no longer available by virtue of the orientation of faces. In the absence of configural processing, face are processed in a piecemeal, featural manner (Bartlett & Searcy, 1993; Searcy & Bartlett, 1996; Stürzel & Spillmann, 2000; Lewis, 2001). The illusion is used to support the hypothesis that we are more sensitive to configural information in upright faces and that there are two modes of processing (configural and featural), with both modes used in upright face processing. Importantly, it is the distinction between the configural and featural processing that is thought to account for the change in appearance.

Evidence supporting this dual-mode account comes from studies examining sensitivity to special positioning. Searcy and Bartlett (1996) had participants perform a same/different task with upright and inverted faces. The faces were manipulated to contain configural (spacing) or featural changes (e.g. blackened teeth). Inversion of the face stimuli was shown to impair encoding (increased the latency of response) of the configurally manipulated stimuli, but not feature manipulated stimuli. Also, Bartlett and Searcy (1993) found that inversion did not reduce the apparent grotesqueness of a face where the grotesque appearance is due to models making a veridical emotional expression. However, it did reduce grotesqueness when the grotesque appearance was due to configural alterations (spacing distortions and Thatcherised faces).

Figure 1.3. The Thatcher illusion (Thompson, 1980). The Thatcherised face on the top right has the eyes and the mouth inverted relative to the face. When this Thatcherised stimulus is inverted (bottom right), the manipulations are not as obvious and the face looks normal. The typical face is presented top left and bottom left for comparison.

The Thatcher illusion has been taken as a measure par excellence of configural processing (e.g. Donnelly & Hadwin, 2003; Rouse, et al., 2004; Maurer et al., 2002) as its perception does not rely on memory, familiarity or the understanding of complex task demands. However, the view that the Thatcher illusion is due to the phenomenology associated with configural processing has recently been challenged. Firstly, Donnelly et al. (2011) asked participants to discriminate Thatcherised from typical faces. The pattern of brain activation plainly showed that the task engages visual processing and socio-emotional processing regions when upright. However, when inverted an extended network of visual processing areas is activated, but there is no activation of socio-emotional brain areas. Therefore, the Thatcher illusion is contingent on more than just areas representing facial configurations for its efficacy; the role of socio-emotional areas is required. The relationship between the Thatcher illusion and emotion perception has also been acknowledged and explored by others (Bartlett and Searcy; 1993; Rothstein, Malach, Hadar, Graif, & Hendler, 2001).

Secondly, recent experiments using measures of processing capacity have shown that the oddity signalled by inversion of eyes and mouths in upright faces does not arise from positive interdependencies between the features; there was no evidence of supercapacity processing for upright Thatcherised faces (Donnelly, Cornes, & Menneer, 2012). Processing capacity refers to the way in which system performance changes as workload is varied. Supercapacity processing is observed when positive interactions between the channels (e.g., between eyes and mouth) cause performance to improve when workload is increased. Supercapacity processing would be the most consistent with configural processing (Wenger & Townsend, 2001) and is therefore expected for upright Thatcherised faces if configural processing is used. For the Thatcher illusion to be a demonstration of configural processing, it requires an unambiguous interpretation of the underlying source of the illusion to be configurality and no other process. Together, these studies show that is not the case.

In summary, there are many open questions in relation to inversion effects, and the interpretation of part-whole, composite faces and the Thatcher illusion studies. The argument presented here is not that the data used to support hypotheses are erroneous; it simply highlights that the data are incomplete and open to different interpretation. It is because of this that the mapping between theoretical frameworks, experiments and data in face processing needs to be explored further.

A New Direction for Face Processing Research

Maurer et al. (2002) provided a review of configural processing, suggesting there are three distinguishable types: (1) detecting first order relations (the spatial layout of features such as eyes above the nose), (2) holistic processing (gestalt properties of features combining together into a perceptual whole) and (3) sensitivity to second order relations (spacing among features). In their view, evidence exists for all three forms of configural processing as distinct from one another (see Maurer et al., 2002). However, as shown previously (see above), there is doubt about how data generated by the tasks outlined above support this theory, and all theories of this type. Methods used to investigate configural processing centred on demonstrating effects consistent with assumptions of the hypotheses, rather than testing the assumptions of the hypotheses directly. This means there is often a problem of circularity in the literature. If configural representation or processing is in force, then a particular empirical regularity (e.g. improved performance in whole relative to part face matching conditions) must be obtained. However, this is sometimes read as implying that if the empirical regularity is obtained then the configural representation or processing must be in force. The existence of a particular empirical regularity need not require that configural representation or processing is functioning, even if it is consistent with it. In this regard, there is uncertainty about the influences on perceptual and decisional processes involved in all tasks thought to reflect configural and/or holistic processing. Therefore, it appears too early to accept any theory that requires certainty in the interpretation of data from the face processing demonstrations outlined. To overcome any potential circularity requires very careful experimentation and analysis to ensure all competing accounts are excluded. From this point forward, unless references is made to a specific hypothesis and therefore the assumptions that go with it, then the term 'configural processing' will be used in its broadest sense to refer to the integration of featural and configural information in face processing, making no assumption of how this information in a face is defined or represented.

Traditionally, study designs for investigating face processing focused on a paradigm that only required a response to one feature, regardless of the number of factors of manipulation in the stimuli (e.g. Farah et al., 1998). Faces are multidimensional forms; therefore, studies looking at face processing should address the simultaneous use of information from multiple dimensions, examining evidence for dependence of face parts in light of theories about configurality. It is clear a new direction for face processing research is required to make progress in understanding the sources of configural processing by directly testing the existing hypotheses in a methodology where responses to multiple stimuli are made.

General Recognition Theory

It is fundamental to understand how dimensions interact in perceptual processing. Therefore, a central term amongst theories is *Perceptual Independence* (PI). The components A and B from the same stimulus are said to be independently perceived if the perception of each is in no way contingent on, or interacts with, the perception of the other. It is also the case if the probability of simultaneously perceiving both A and B is equal to the probability of perceiving A times the probability of perceiving B (Ashby & Townsend, 1986). A problem facing researchers is that perceptions are not directly observable and pass through a decisional process that selects the most appropriate response for the situation. Therefore, decisional processes fundamentally alter perceptions making it difficult to test for PI. Ashby and Townsend (1986) developed *general recognition theory* (GRT) to encompass the range of terms and mechanisms for independence criteria that arose from the difficulty to test for PI. A stronger test of PI was produced and a resolution was found to the confusion created by the existence of many tests of PI. GRT also encompasses decisional processes which were missing previously. GRT is a multidimensional extension of *signal detection theory* (SDT), a method of analysis that addresses perceptual and decisional effects separately (see Appendix A for an account of SDT and its value). GRT provides formal definitions for the ways in which dimensions can interact during recognition, and thus provide a framework for modelling the dependencies that are indicative of configural processing. GRT has been pinnacle in a new approach to understanding configural processing.

GRT examines relationships between different dimensions on multiple axes. It assumes a probability within perception, meaning that perception can vary from one trial to the next and, therefore, represents stimuli as multivariate probability distributions. A theoretical model of two stimulus dimensions, each with two levels, would have four distributions across two axes and two response criteria (see Figure 1.4).

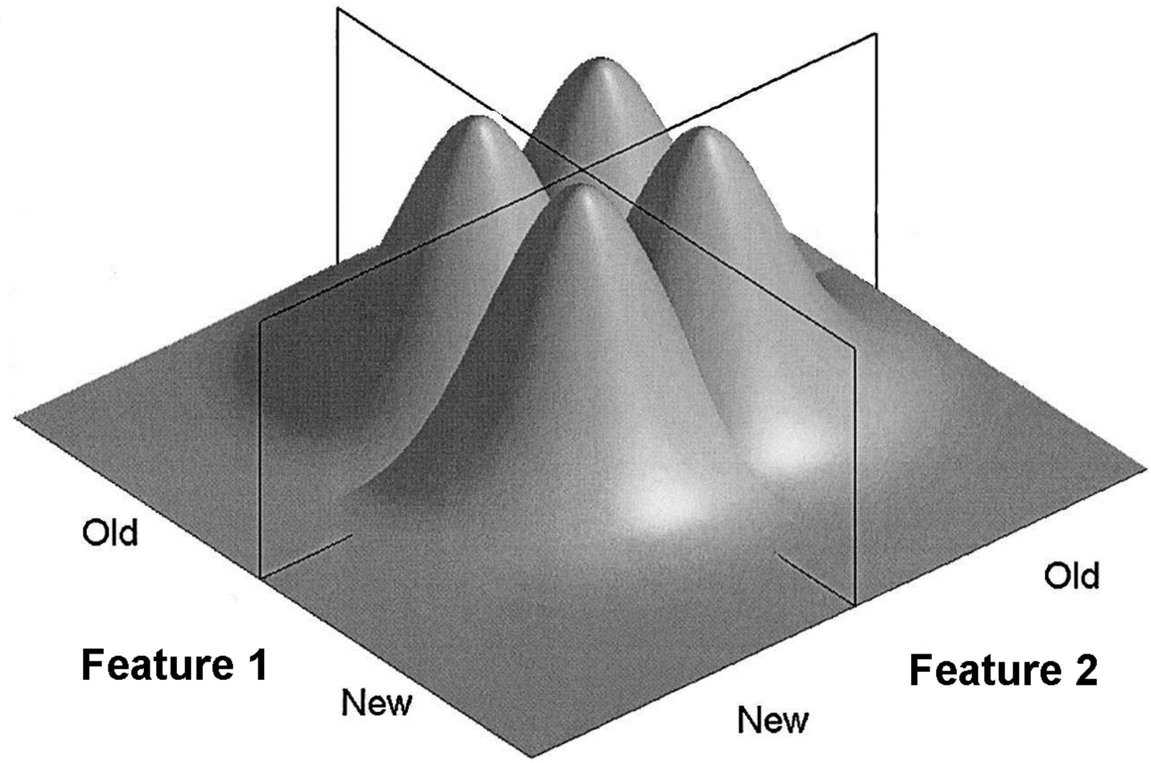


Figure 1.4. Diagram showing a GRT distribution of equal likelihood. The distribution represents a factorial design from a feature recognition task. There are two features taking one of two possible levels: old or new. Diagram modified from Wenger and Ingvalson (2003).

Figure 1.4 illustrates the distributions for four stimuli, defined by two dimensions which represent two features in a face. It is a factorial design with each feature taking one of two possible levels: old or new. Dimensional interactions in multidimensional stimuli can be characterized by either perceptual or decisional factors. There are three origins of these effects: (1) Violations of perceptual independence (PI), (2) violations of perceptual separability (PS) and (3) violations of decisional separability (DS).

Figure 1.5a shows the equal likelihood distribution seen in Figure 1.4, but seen from above and cut horizontally across the plane to show contours of equal likelihood. Each circle represents the multivariate normal distribution for each stimulus. There are no interactions in the stimuli. Perceptual independence holds when the perceptual effect of one dimension is statistically independent of the perceptual effect of another dimension (the variability in the perceptions are uncorrelated). Figure 1.5b represents violation of PI in distribution A_1B_1. This demonstrates a bivariate correlation within the perceptual distribution (the oval). Such correlations can only occur under multiple dimensions, and therefore this concept does not exist for traditional one-dimensional SDT. Only violation of PI is a within-stimulus effect as it can be observed in a single stimulus (distribution) making it the evidence for configurality that is most consistent with previous accounts.

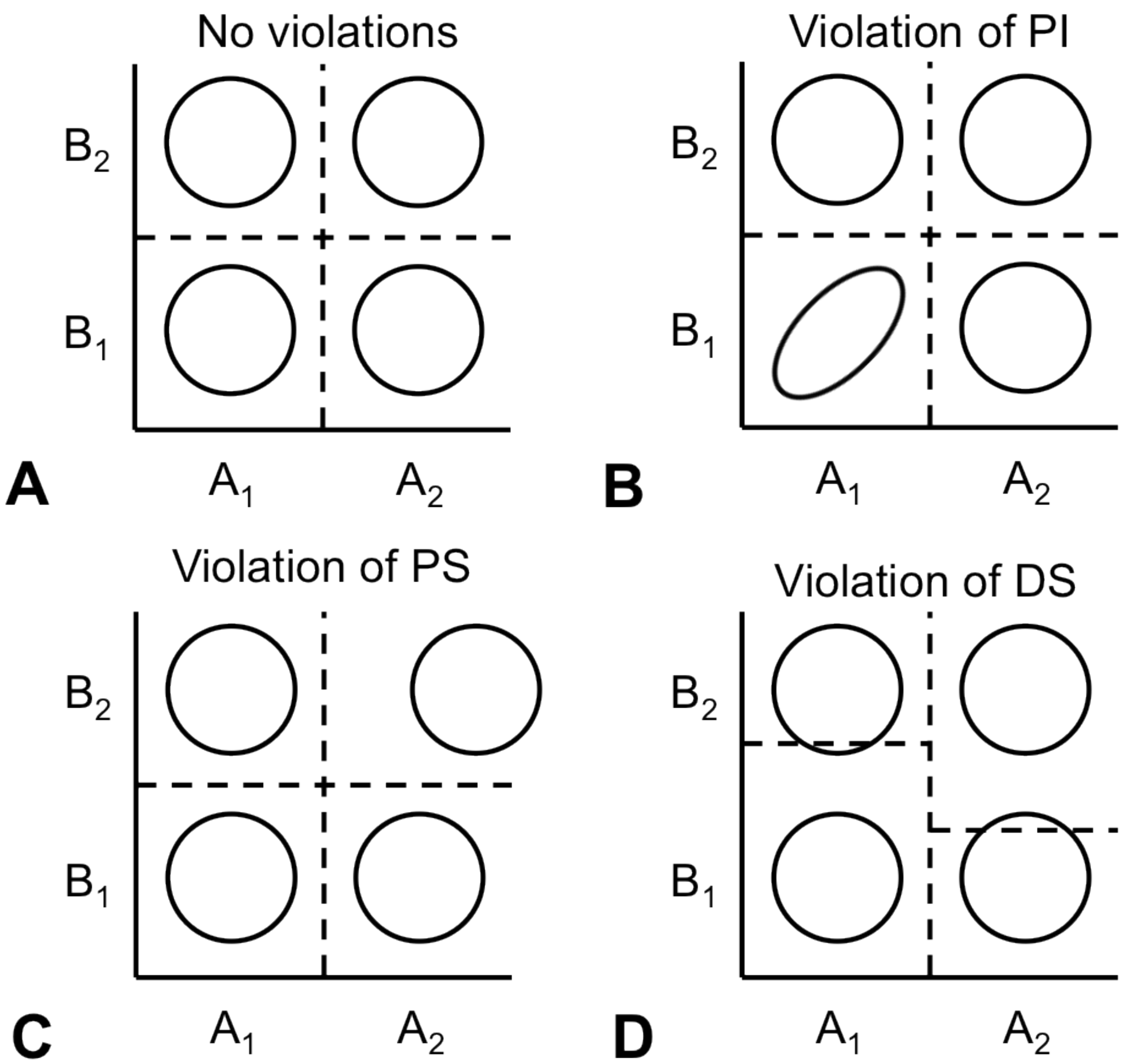


Figure 1.5. Diagrams showing GRT distributions cut through an arbitrary plane and viewed from above for (A) distributions of equal likelihood and no violations, (B) violations of PI in one distribution, (C) distributions with a violation of PS and (D) distributions with a violation of DS.

When PS holds, perception of one dimension is unaffected by the level of the other dimension. When PS is violated, the perception of one dimension depends on the level of the other dimension. This measure uses the d's which are the sensitivity measures for the distributions. Violation of PS is represented in the non-rectangular arrangement of perceptual distributions (Figure 1.5c), with a greater *d'* in level two of feature B compared to level one. A violation of PS is different to a violation of PI as it is based on unequal distances between distributions (measured as *d'*) rather than correlated perceptions within one distribution (PI).

Decisional separability holds when the decision about one dimension is unaffected by the level of the other dimension. The boundary used for decisions about one dimension is in the same location irrespective of the level of the other dimension. When DS is violated, the response about one dimension depends on the level of the other dimension. This measure uses the criterion (bias values) about responding to the distributions. When DS is violated, the location of the decision boundary for one feature will not be the same across levels of the other feature. The decision boundary will shift (Figure 1.5d). The decision bound for feature B in this example is a piecewise step function between one level and the next, but the decision bound could also change continuously as a diagonal line to show the change in decision bound across the levels. Violation of DS indicates a decisional component, not perceptual components like violations of PS and PI.

General recognition theory provides an advantageous method as it can represent encoded information from multiple dimensions in a factorial design. This allows simultaneous responses about multiple attributes as well as perceptual and decisional effects to be separated. The *feature complete factorial design* (e.g. Kadlec & Townsend, 1992) involves constructing stimuli in factorial combinations so that each possible level on one dimension is combined with each possible level on the other. A *complete identification task,* in which participants identify the level of each dimension, is advantageous as it can define conditions under which manipulations to one feature of a face might influence responses to another feature. The correct responses are recorded along with the types of incorrect response in a confusion matrix. A confusion matrix of responses allows inferences to be made from the context of shifts in levels of performance. In contrast, in an incomplete design, these measures cannot be examined.

Kadlec and co-workers (Kadlec, 1995,1999; Kadlec & Hicks, 1998) presented a multidimensional signal detection analysis (MSDA) program grounded in the principles of GRT allowing empirical investigation of the perceptual and decisional dependencies. Since these papers were published, the calculations have developed and become more reliable to include converging techniques for inferring dependencies (e.g. Cornes, Donnelly, Godwin, & Wenger, 2011). The theory, the design of the complete identification paradigm and the method of running the analysis (MSDA) are all linked in a well formed methodology to create a GRT framework for exploring configural face processing. The measures used for MSDA analysis and the evidence required for inferences to be made are outlined in Chapter 2. These methods in particular have been applied to research on face processing but there are other methods for analysing GRT data (e.g. Macho, 2007; Thomas, 2001; DeCarlo, 2003) and the probit model approach (DeCarlo, 2003) is described in more detail in Chapter 2.

GRT Studies in the Face Processing Literature

Recent studies have used GRT to re-examine the nature of classic demonstrations of configural processing that have been dominant in the face processing literature. Wenger and Ingvalson (2002) used constructs of GRT to investigate the scale of configural processing for internal representation of a face. The simultaneous perception and memory for more than one dimension for a face was required in the task, which allowed identification of the type of processing used. The Farah et al. (1998) study provided accuracy results in their manuscript, allowing the data to be re-analysed using signal detection. Wenger and Ingvalson (2002) re-examined the results of Farah et al. (1998) Experiment 1 and found a shift to a more conservative response criterion when the irrelevant feature was incompatible rather than compatible. This suggests a decisional process is contributing to the behavioural effect of holism demonstrated in the study. It also revealed that using evidence from the study to support the holistic perception hypothesis actually contradicts the assumption of a purely perceptual effect. Wenger and Ingvalson (2002) aimed to test the assumptions of the holistic perception hypothesis directly by testing for violations of the three central constructs of GRT (PI, PS and DS). They suggested violations of any of these constructs of GRT could produce performance consistent with the behavioural marker of the holistic perception hypothesis. However, only violations of PI or PS would support the idea of a holistic internal representation. Stimuli were presented upright as this is the optimal orientation for holistic encoding (Farah et al., 1998).

The task used four classes of stimuli: faces, animals, vehicles and geometric shapes. The eyes and nose, and the equivalent top and middle features in the control stimuli were manipulated simultaneously in a complete factorial design. Therefore, four versions of each 'original' stimulus were created: no changes were made, the top features (eyes or headlights) were enlarged; the middle feature (nose or number plate) was enlarged; or both the top and middle features were enlarged. Participants were shown a study image and asked to respond to the test image that followed using one of four responses. These were no changes; change to top feature but not middle; change to middle feature but not top; or changes to both features. Wenger and Ingvalson (2002) also manipulated the retention interval (RI) between the study face and the test face using a backwards counting task. Either the RI was 0 (no backwards counting task), 3, 9 or 15 seconds. The authors wanted to see if there was any evidence for holism as a result of retention (non-zero RI) and if there were any violations found at the 0 second RI.

The responses were examined for hit rates and bias and were found to be analogous to the effects of sensitivity and bias in Farah et al. (1998) Experiment 1. Increased hit rate to features was found when the status of the other feature was 'old' (the same) compared to when it was 'new'. Also, a conservative criterion shift was observed when the status of the other feature was 'new' (not the same). In Wenger and Ingvalson's (2002) second experiment with inverted stimuli they found the same effect of hit rate observed in Experiment 1 was present but attenuated, and the conservative shift in bias was still present. Experiment 3 encouraged participants to encode the stimuli as meaningful wholes in an attempt to enhance holistic processing. Performance for one feature was affected by old or new status of the other feature when a face was presented upright, consistent with predictions of the holistic perception hypothesis. This suggests the new method is measuring the same principles as previous methods.

For Experiment 1, the GRT analysis revealed only a consistent violation of DS across face stimuli, but these violations were observed at all RIs. Only limited evidence was found for violation of PS and no evidence for violation of PI. A violation of DS alone suggests the effects supporting holism are based in the decisional criteria about the dimensions, not the perceptual representation. Therefore, results were not consistent with the holistic perception hypothesis, but instead suggest configural processing could have a decisional component. The violation of DS was evident across all stimulus types, when stimuli were inverted (Experiment 2), and when stimuli were paired with an adjective of either positive of negative valence (Experiment 3). No evidence for a violation of PI was found. Violation of PI can be thought of as evidence for the strongest form of holism as, unlike PS and DS, it suggests correlated perceptual effects that are specific to a single type of face. As inconsistent evidence was found for PS in study one, Wenger and Ingvalson (2003) replicated the experiment using a different test of PS by examining the diagonal *d'*s and again found the consistent violations of DS only.

A problem of moving toward a new methodology and analytical framework was the inability to directly compare the findings to previous studies. GRT inferences can only be made in a complete identification paradigm; a methodology which requires divided attention across the whole stimuli. Previous studies required focused attention to one feature, regardless of multiple changes to the stimuli. Therefore, this posed a problem to interpretation of the Wenger and Ingvalson studies (2002, 2003) to the wider face literature as the findings could be due to a differential application of attention in the tasks.

Richler et al. (2008) were able to provide this comparison by applying a new task procedure to the domain. The *sequential matching task* provided a direct comparison of results from a selective attention method (response to a single target feature), with a complete identification method (a single response made about the status of two features, as used by Wenger & Ingvalson, 2002, 2003) by encompassing features of both methods. All three tasks were used in Richler et al.'s study. In the sequential matching task, a response is made to each of two features. In the selective attention task, one judgement was made about whether the top half of a test face is the same or different as the top half in the study face shown previously. This was compared to the first response in the sequential matching task. The complete identification task required one response based on status of both features: both halves same, top half different bottom half same, bottom half different top half same or both halves different. Comparison could then be made to results of both judgements made in the sequential matching task.

Participants in Experiment 1a of Richler et al. (2008) completed a selective attention task and a sequential responses task. Participants in Experiment 1b undertook a complete identification task and a sequential responses task. The congruency effect for sensitivity was found when comparing individual responses from the sequential responses task to the single response in the selective attention task (higher sensitivity if the non-target face half was the same status as the target half). This replicated behavioural evidence for holism (Farah et al., 1998). Also, both the complete identification task and sequential responses task revealed an effect of status of the other feature, where sensitivity was higher and bias was more conservative when the other feature was the same rather than different. This was shown to be consistent with the pattern in the selective attention task. The two feature complete designs also revealed violations of DS. Unlike Wenger and Ingvalson (2002, 2003), Richler et al. (2008) also found violations of PS but, consistent with Wenger and Ingvalson (2002, 2003), no violations of PI were found.

The systematic misalignment of a composite face is suggested to reduce the effect of holism (Young et al., 1987). This was also of interest as Richler et al. (2008) questioned whether systematic misalignment would correspond to changes in GRT inferences. They found that the congruency effect decreased with misalignment. A consistent violation of PS was found for aligned faces, but to a lesser degree for misaligned. Decisional separability was violated across all misaligned conditions, and was significantly affected by the extent of misalignment. This suggested that changes in the congruency effect are linked to a decisional component. There may also be a perceptual component, and importantly, the difference between alignment and misalignment was underpinned by decisional factors.

Richler et al. (2008) provided a means to successfully compare two previously incomparable methods, revealing the violations of DS were not due to the task demands of a two feature paradigm. They have provided validity of the new sequential response methodology as the GRT results are replicated from the complete identification method; thus, they are not due to dividing attention across the stimuli. Also, they suggest that violations of PI, the construct that most strongly represents holistic processing, are not contributing to the behavioural effect of configurality in an aligned composite face.

Cornes (2011) investigated the nature of the Thatcher illusion (Thompson, 1980), examining whether dependencies exist between features in a face and across orientation by using feature inversion manipulations. Participants were presented with ‘normal’ and ‘Thatcherised’ faces and church control stimuli in which the eyes (windows), the mouth (door), or both the eyes and mouth were inverted. The stimuli were also presented either upright or inverted. Participants had to respond to three questions: (1) whether the outline was upright or inverted, (2) whether the eyes (windows) were upright or inverted and (3) whether the mouth (door) was upright or inverted. Participants also had to complete grotesqueness ratings for each of the stimuli.

Cornes et al. (2011) created multiple GRT models that indicate presence of the Thatcher illusion, allowing predictions about the many possible outcomes of GRT inferences. This provided a way of conceptualising how different GRT violations and their combinations could produce the effect. The models indicate that violation of PI alone cannot produce evidence for configurality in the Thatcher illusion, but violations of PS and/or DS could. This is consistent with simulation results presented by Richler et al. (2008) that also suggested violation of PI alone cannot generate evidence for configurality observed in the composite face effect. These results also highlight that there may be multiple origins of configurality.

The results of Cornes et al. (2011) demonstrated ratings of grotesqueness were higher for fully Thatcherised stimuli in an upright rather than inverted orientation. However, there were no qualitative differences between grotesqueness ratings of church and face stimuli. For the MSDA, violations of PS and DS were found that differentiate the inverted from upright faces. However, no within-stimulus configurality (violation of PI) was found. This is consistent with the simulated models and previous studies that demonstrate a decisional locus for configural effects is present. For configurality in the traditional sense, DS is expected to hold as dependencies are expected to be perceptual, and holism is expected to reside within a stimulus (violation of PI). In these three examples, evidence is only present for dependencies across stimuli (violation of PS). Correlations between the MSDA statistics and the grotesqueness ratings support both perceptual and decisional effects in the Thatcher illusion. Also, as expected, there were individual differences. Therefore, criteria that underpin the Thatcher illusion show evidence for individual differences (Cornes et al., 2011). This is in line with the simulations that show the Thatcher illusion can be produced by multiple means.

These findings raised some questions about how the Thatcher illusion can be explained if it is not associated with perceptual configurality within a single face. The findings of Cornes et al. (2011) suggest that the effect of orientation is not face specific, and the difference in results for the inverted stimuli may be due to a quantitative rather than qualitative change in the type of processing used for inverted compared to upright faces.

Violations of DS are present in each of the outlined complete identification face processing tasks. Violations of PS are revealed in two of them but violations of PI are not revealed in any of them. All three tasks (Wenger & Ingvalson, 2002, 2003; Richler et al. 2008; Cornes et al. 2011) were analysed using similar MSDA computations. Replication of results with a different method of GRT analysis would validate these findings.

Alternative Methodologies Used to Explore Configurality

There have been attempts to address the nature of configurality using other novel approaches and methodologies, such as information-processing which examines architecture, stopping rule and dependency. Predictions can be formed for each of these measures based on theories of face processing. Architecture is examined to see whether features are processed simultaneously (parallel), sequentially (serial) or pooled into a single perceptual unit (coactive). The stopping rule that participants use to decide whether faces are, for example, the same or different can either be self-terminating (where search can terminate on a single feature) or exhaustive (where all features must be processed). Dependency between processes is investigated to see whether the processing of one type of information has an effect (dependency) or not (independence) on the rate at which another type of information is processed, and whether dependency facilitates or inhibits the overall decision.

Fifíc and Townsend (2010) examined failure of selective attention in faces with spacing manipulation as a test of the holistic processing hypothesis. Despite failure of selective attention being necessary for holistic processing, both analytic and holistic models of information processing measures were shown to predict it. This suggested evidence for configural processing can be accounted for by non-configural models. Fifíc and Townsend (2010) also suggested that slower RT when a face context changes may be related to the adjustment of a decisional criteria. Ingvalson and Wenger (2005) tested assumptions of the dual-mode hypothesis by relating them to information processing predictions. They examined faces and geometric faces with colour and spacing manipulations. They were able to find evidence consistent with some information processing predictions of the dual-mode hypothesis, but failed to find evidence of independence of featural and configural sources of processing. Processing capacity can also be examined, which is closely related to dependency. It refers to the way in which system performance changes as workload is varied and can be considered limited capacity, unlimited capacity or supercapacity processing (see Donnelly et al., 2012). Donnelly et al. (2012) showed the oddity signalled by inversion of eyes and mouths in upright Thatcherised faces does not arise from positive interdependencies between the features (supercapacity processing).

These methodologies allow a direct test of the assumptions of configurality in order to re-address behavioural results that are often used in place of testing hypotheses. It is a converging approach to the GRT framework and provides added rationale to re-examine previous demonstrations of configurality as inconsistencies with face processing hypotheses are revealed. In addition to testing for dependencies indicative of configurality, GRT as a framework shows whether the dependencies are perceptual or decisional in nature.

Rationale for Research Project

Previous research examining dependencies between features was based upon demonstrations that have been shown to have multiple interpretations. This has led to difficulties in discriminating between competing accounts of face processing. Absence of a precise definition has resulted in previous demonstrations of configural processing now being accounted for by decisional rather than perceptual processes (Wenger & Ingvalson, 2002, 2003; Richler et al. 2008; Cornes et al. 2011). However, a lack of detailed understanding about the nature of configural information remains despite attempts to define configural processing in functional, mathematical and neural terms.

Distinctions are essential regarding the way in which configural processing of a face might be produced. GRT analysis can provide these distinctions, along with a strong test of hypotheses, by defining configural processing mathematically in terms of the three constructs: violations of PI, PS and DS. It can also provide a test of alternative hypotheses where previous tests have failed. A framework for re-analysis of configural processing and the design of experiments has been provided, so should continue to be exploited. This is fundamental to an improved methodology as results from previous studies which adopt an incomplete approach (e.g. the partial design for composite face studies) are difficult to reconcile with findings from feature-complete designs.

Chapter 2 provides a comparison of the GRT inferences in relation to behavioural effects in three face processing tasks used to signal configural processing. Also, it examines whether there are different types of configural processing demonstrated in different behavioural tasks, and whether the evidence revealed in the violations matches the theory for which these tasks are suggested to support. It also compares the inferences from two different approaches for analysing the results in the GRT framework in order to investigate why violations of PI may not have been observed in previous studies. The value of GRT and the probit method are demonstrated in relation to the issue of identifying sources of configurality in face processing. The task with the best evidence for configurality (Thatcher illusion) is then explored further with converging evidence from the event-related potential (ERP) technique.

Recent progress in the methodology and analysis of electrocephalogram (EEG) has revealed that effects of face processing tasks on ERPs may be due to the poor control of low level visual properties across the face and control stimuli (Rouselett, Husk, Bennett & Sekuler, 2008). Although the ERP paradigm is well established in the face processing literature, procedures must be adapted to ensure effects are not simply artefacts of poorly controlled stimuli. Also, using a data driven approach rather than a reliance on previous demonstrations will prevent ERP studies focusing on known perceptual markers. Chapter 3 addresses these criticisms and examines physiological markers for configural processing in the Thatcher illusion using an ERP paradigm. Potential mapping between ERP correlates and the violations revealed for the Feature Orientation Task (Thatcher illusion task) in Chapter 2 is discussed.

In the following chapters, the GRT framework is applied to developmental and special populations, both of which can provide valuable evidence about face processing. Within the developmental domain, discrimination between perceptual and decisional sources of configural effects across age can be determined. Cornes et al. (2011) suggested that the mechanisms that increase sensitivity and produce shifts in response criteria across development should be examined to gain a better understanding of the development of face processing. Whether developmental effects are due to quantitative or qualitative changes in either decisional and/or perceptual processes can be examined. Chapter 4 tests the feasibility of a GRT Thatcher illusion task applied to a developmental study. There are many challenges for applying the methodology to young participants. The tasks traditionally require large numbers of trials, and the tasks themselves can be very difficult. Therefore, methods to reliably collect data from children in complete factorial paradigms must be developed.

Similar issues arise when applying the GRT methodology to specialist populations, specifically for prosopagnosia. When the condition is acquired, there is often diffuse brain damage which can make tasks requiring a memory component very challenging. In addition, aggregating data over the group of participants is not appropriate as it requires the assumption that patients are homogeneous with respect to the nature of their deficits (Shallice, 1990). Progress to a methodology appropriate for such participants is required. In Chapter 5, the suitability of the GRT framework applied to an individual with acquired apperceptive prosopagnosia is explored. There is limited research of the Thatcher illusion and prosopagnosia. Thus, the basic behavioural evidence in relation to the illusion needed to be obtained. Sensitivity to the Thatcher illusion is assessed alongside emotional face processing in light of recent evidence pointing to socio-emotional activation in relation to the illusion (Donnelly et al., 2011).

Throughout this thesis, the value of applying the GRT framework to the issues relating to poor understanding of the underlying nature of configurality is presented. GRT provides a sophisticated account of the data by conceptualising configurality as perceptual and decisional dependences. However, it also allows robust replication of previous behavioural effects of face processing such as the inversion effect and sensitivity to the Thatcher illusion.

Research Aims

(1) To provide a comparison of GRT inferences across different tasks measuring behavioural effects of configural processing in order to examine whether there are different routes to configurality and whether the tasks use the same underlying concepts (violations of PI, PS and DS). Use GRT to directly test the assumptions of configurality in these tasks.

(2) To explore converging evidence for GRT findings with frameworks that pursue a well-controlled and an unbiased approach (is open to testing for non-perceptual mechanisms).

(3) To apply the GRT framework to a developmental population and an individual with prosopagnosia to determine what these populations can reveal about sources of configural face processing.

Chapter 2

Identifying Sources of Configurality in Three Face Processing Tasks

In Chapter 1, I described how many studies on face perception thought to indicate configural face processing are subject to different interpretations. Furthermore, I showed how consideration of configurality in face processing through the lens of general recognition theory (GRT) might be helpful. In particular, the typical behavioural markers of configurality might be contributed to by both perceptual and decisional factors. In this chapter, this theme is continued by exploring the performance of participants across variants of three classic tasks thought to demonstrate configural face processing. In each case, the tasks were setup as complete identification tasks, and the data has been analysed within the terms of GRT. The study in Chapter 2 has been published in Frontiers in Psychology (Mestry, Menneer, Wenger, & Donnelly, 2012).

One of the key issues for research on configurality in face perception is whether some commonality exists across the effects reported with different phenomena. Studies of face perception have used various tasks to explore how faces are processed as either wholes or configurations. Amongst the most common examples are the holistic effects in part-whole studies (Tanaka & Farah, 1993; Davidoff & Donnelly, 1990) and composite face tasks (Young, Hellawell, & Hay, 1987). In addition, a face specific effect known as the Thatcher illusion (Thompson, 1980) is often used to mark the presence of configural processing (Bartlett & Searcy, 1993; Maurer, Le Grand, & Mondloch, 2002; Donnelly & Hadwin, 2003).

Typically, two conclusions are made in the face literature: upright faces are processed holistically or as configurations. By the holistic account, the perception of whole faces occurs automatically and at cost to the perception of face parts (Tanaka & Farah, 1993; Davidoff & Donnelly, 1990). By the configural account, second-order relationships are formed between features (Diamond & Carey, 1986), and inverted faces are processed with effort and in a piecemeal fashion as features. These conclusions are based on empirical signatures of configural face processing and are used as evidence for hypotheses in the literature. However, to overcome any potential circularity as outlined in Chapter 1, assumptions of hypotheses must be tested directly.

One approach is to use theoretically grounded formal definitions of configural representation or processing (e.g. Ashby & Townsend, 1986; O'Toole, Wenger, & Townsend, 2001; Townsend & Nozawa, 1995). Use of these formal definitions allows mapping between tasks and theories in terms of a set of mediating constructs. The basis for this work is the set of theoretical definitions of configural representation provided by GRT (Ashby & Townsend, 1986). In order to link these theoretical definitions to data, the feature-complete identification paradigm is adopted (e.g., Ashby & Townsend, 1986; Kadlec & Townsend, 1992b; Kadlec & Hicks, 1998; Townsend, Hu, & Ashby, 1981). In this paradigm, separate responses are required for each of the features that can vary across all trials. These data can then be analysed in two ways. Firstly, using group-based (aggregate) analyses of signal detection measures to determine whether there are differences in sensitivity and bias, and whether any differences are consistent across conditions. This is considered the norm for evaluating behavioural data in visual cognition experiments. Secondly, analysis of multidimensional measures of sensitivity and bias in order to draw inferences regarding independence and separability as defined within GRT (Ashby & Townsend, 1986). This second set of analyses allows inferences to be made at the level of the individual, and in terms specified by GRT.

As outlined previously, within a feature complete identification paradigm, the evidence for configurality comes from the perception of, or responses to, one feature (e.g. eyes) being shown to be dependent on the status of other features (e.g. mouth). Encoded dimensions (e.g. eyes and mouth) may interact perceptually, either at the level of the individual stimulus or across the set of stimuli, and may also interact decisionally in the generation of a response. Perceptual interactions are characterised via the shape and locations of the distributions of perceptual evidence which arise from each stimulus type (e.g. eyes-upright, mouth-inverted). These perceptual interactions are represented in GRT by violations of perceptual independence (PI) and/or perceptual separability (PS). Decisional interactions are characterised via the shape and location of decision bounds between the distributions, and are represented in GRT as violations of decisional separability (DS).

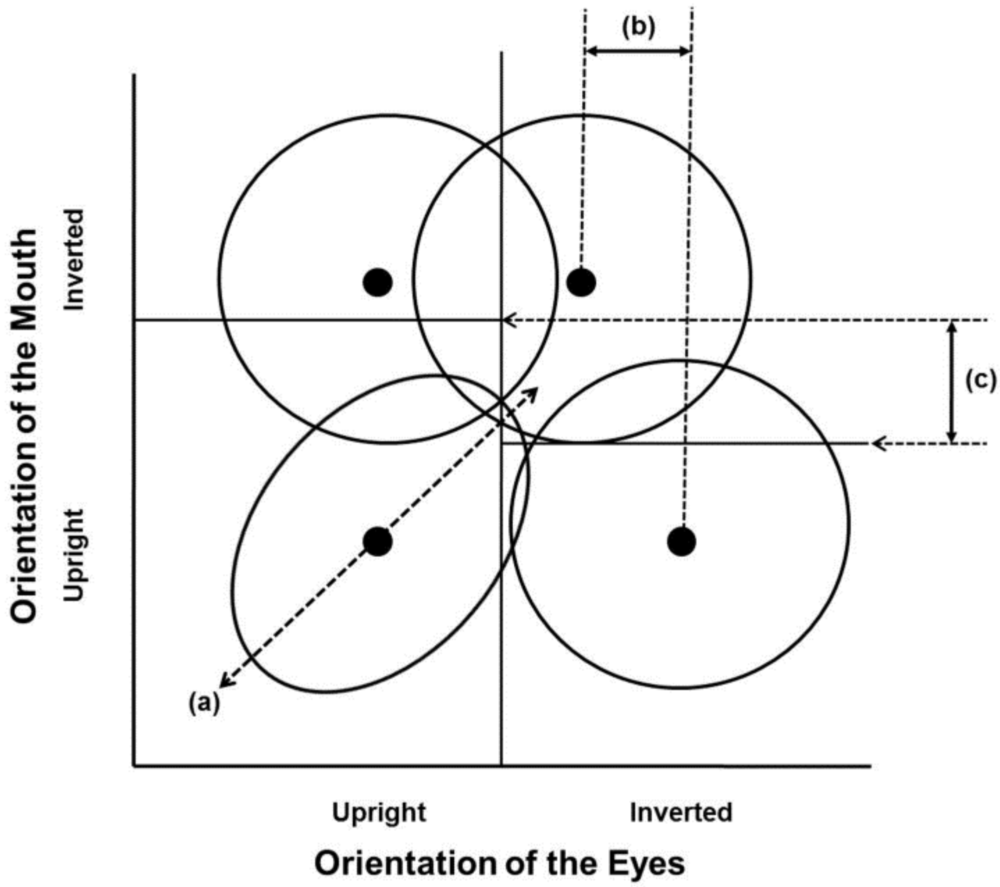


Figure 2.1. Schematic representation of the three ways of theoretically representing configurality in GRT. (A) illustrates the correlation within the perceptual information predicted by a violation of perceptual independence. (B) illustrates the differences in the marginal means for the orientation of the eyes predicted by a violation of perceptual separability. (C) illustrates the differences in the location of the decision bounds for the orientation of the mouth as a function of the level of the eyes predicted by a violation of decisional separability. Figure from Mestry, Menneer et al. (2012).

Violations of PI, PS and DS can be illustrated schematically. For example, if one dimension of the stimulus is the eyes with levels upright and inverted, and the other dimension is the mouth with levels upright and inverted, then within this two-by-two framework, four stimulus types can be represented by four bivariate probability distributions; one for each combination of the two levels of each dimension. Figure 2.1 shows these bivariate distributions as four contours. A violation of PI in the upright-upright stimulus could be represented by a positive correlation in the bivariate distribution of perceptual information (Figure 2.1a). This correlation would represent a within-stimulus interaction between the perceptual information for the state of the eyes and mouth. A violation of PS (Figure 2.1b) would be represented by a shift in the marginal means for one or more levels of each of the two dimensions, such that overall sensitivity to the orientation of the eyes is greater when the mouth is upright than when it is inverted. A violation of DS (Figure 2.1c) would be represented by a shift in the location of the decision bounds that separate the evidence space into response regions, such that when the eyes are upright participants are more likely to respond mouth-upright than when the eyes are inverted. In this way, GRT provides three ways of theoretically defining how features can interact, such that the percept or response to one feature is dependent on the status (or level) of the other feature. Therefore, GRT provides three ways of theoretically representing configurality.

Formally, a GRT model relies on the parameters of the probability distributions and the extent to which the perceptual evidence supports different responses. In the example previously discussed (and represented in Figure 2.1), if an assumption is made that the distributions are bivariate Normal then each bivariate distribution is completely specified by a vector of means and a covariance matrix (Equation 1):

$$\mu = \begin{bmatrix} \mu_E \\ \mu_M \end{bmatrix}, \Sigma_i = \begin{bmatrix} \sigma_E^2 & \rho\sigma_E\sigma_M \\ \rho\sigma_E\sigma_M & \sigma_M^2 \end{bmatrix} \qquad (1)$$

Where μ is the mean, σ is the standard deviation and ρ is the correlation. E indicates the eye dimension and M indicates the mouth dimension.

If it is then assumed that the decision bounds are linear, a complete GRT model of hypotheses for configurality is given by four mean vectors and covariance matrices and two or more decision bounds. A violation of PI for any one stimulus is defined as a non-zero value for the correlation parameter ρ in the covariance matrix for that stimulus. In the example (Figure 2.1a), ρ for the upright/upright stimulus would be positive. A violation of PS for one of the stimulus dimensions is defined as a difference in the location parameters (the ρs for that dimension in two of the covariance matrices), a difference in the variability parameters (the ρs for that dimension in two of the covariance matrices), or both, for one dimension changing across levels of another dimension. In the example (Figure 2.1b), the marginal mean for eye orientation when the mouth is upright is to the right of the corresponding marginal mean when the mouth is inverted. Finally, a violation of DS is defined as the criterion value(s) for assigning responses in one dimension changing across levels of another dimension. In the example (Figure 2.1c), the location of the decision bound for the orientation of the mouth when the eyes are upright is above the location of the corresponding decision bound when the eyes are inverted. It is important to realise that any of these violations (or their combinations) could lead to differences in accuracy or reaction time (RT) across the conditions of classic face processing tasks (e.g. recognition across upright versus inverted faces, aligned versus misaligned faces, whole versus part faces), used to support inferences about configural processing.

GRT in the Face Processing Literature

As described in Chapter 1, several studies have used GRT to explore holistic effects (Wenger & Ingvalson, 2002, 2003), composite effects (Richler, Gauthier, Wenger, & Palmeri, 2008) and the Thatcher illusion (Cornes, Donnelly, Godwin, & Wenger, 2011). All of these studies have used factorial paradigms.

Wenger and Ingvalson (2002) used an odd-new recognition paradigm for the size of two features within a stimulus. Size changes could be made to two possible target features within faces and non-face objects and responses were required for both. Retention interval (RI) was varied and the experiment was also repeated with inverted stimuli (Experiment 2). Wenger and Ingvalson (2002) found violations of DS across all stimulus types (upright and inverted) and differing RIs in Experiments 1-3. Very limited evidence was found for violations of PS at longer RIs for upright faces in Experiment 1. Evidence suggested a shift to a relatively more conservative response criterion when the other feature was new (different) rather than old (the same) for the upright faces.

Richler et al. (2008) examined the composite face effect and demonstrated congruency effects can be seen in the traditional task where participants only had to respond to the target face half (selective attention task) and complete factorial designs where two responses are required. For the single responses, Richler et al. (2008) found a congruency effect; sensitivity was higher when the non-target face part had the same status as the target face half. This effect was reduced with misalignment. When responses were required to both face halves (either made sequentially or simultaneously) effects of congruency were also found. When the status of the other face half was ‘the same’ as previously shown, sensitivity was higher for upright composite faces. This effect was reduced as the composite faces were misaligned. Also, there was a shift to a relatively more conservative response bias when the status of the other face half was 'different' compared to when it was 'the same'. This means participants were more likely to say the target face half was the same if the non-target face half was the same. This shift was reduced by misalignment. Richler et al. (2008) found violations of PS and DS for the task and noted that a reduction in the congruency effect with misalignment was coupled with a reduction in evidence for violations of DS (reduced differences between marginal *c* values).

Cornes et al. (2011) used Thatcherised faces and churches in a three-response paradigm as they required responses to two internal features and the outline. They noted that perception and judgement of one of the internal elements was affected by the external form more than by the level of the other internal element. Violations of PS and DS were documented, but reliable violations of PI were not found (Cornes et al., 2011). Overall, more violations of PS were found compared to DS, and more violations were observed when the stimuli (faces and houses) were inverted. Also, large individual differences within the data were reported, suggesting the Thatcher illusion can be generated in more than one way.

In summary, whereas all three studies reported violations of DS, and two showed violations of PS, none showed violations of PI. These results reveal both similarities and dissimilarities across paradigms. Any inconsistencies could occur because of genuine differences between the processes required to perform tasks accurately. Alternatively, they could arise through the use of different groups of participants or some other trivial difference between tasks (such as different stimulus sets). Therefore, in the present study, a single group of participants was tested across the tasks to compare the findings from the GRT analyses.

Multiple Methods for GRT

The data obtained in feature-complete factorial designs are typically summarised in an identification/confusion matrix. Two methods of analysing these data have been used in order to draw inferences regarding whether there are any violations of PI, PS, or DS. The first of these is the oldest within the tradition of work with GRT. It involves a set of parametric and non-parametric comparisons, and is sometimes referred to collectively as multidimensional signal detection analyses (MSDA). The second involves one or more methods of directly estimating the parameters of the underlying multivariate Normal distributions and is referred to as the probit model. Although these two approaches have not always been used together, they provide two potential sources of converging evidence whose dual use is advantageous given long-standing concerns regarding potential inferential problems (beginning in Ashby & Townsend, 1986).

As outlined previously, many of the applications of GRT to questions regarding the perception of, and memory for, faces have reported violations of PS and DS but rarely violations of PI (Wenger & Ingvalson, 2002; 2003; Richler et al., 2008; Cornes, et al., 2011). On the basis of these results, one might conclude that evidence of configurality in upright faces is driven by shifts in perceptual sensitivity and bias for features in upright relative to inverted faces. This conclusion is at least superficially incongruent with the vernacular conception of configural processing, which anticipates dependencies within a given stimulus (i.e. violation of PI) as well as relationships between stimuli (i.e. violations of PS and DS). Therefore, this conclusion deserves additional scrutiny.

In the present study, this failure to observe violations of PI is examined using a novel adaptation of a statistical method for estimating the parameters of the underlying multivariate distributions (DeCarlo, 2003). Preliminary work with this approach has suggested that it may have greater sensitivity to the presence of non-zero correlations than has been reported with other methods (Menneer, Wenger, & Blaha, 2010, 2012; Menneer, Silbert, Cornes, Wenger, Townsend, & Donnelly, 2009). The approach uses multiple probit models to directly estimate the parameters of the underlying multivariate Normal distributions (DeCarlo, 2003). By allowing direct estimation of the correlation parameter for each multivariate distribution, there may be a greater chance of detecting violations of PI in upright faces than has previously been the case. It should be noted that methods used to date have intentionally been conservative with respect to inferring violations of any of the constructs. This method of analysis is different to that used in the previous GRT tasks examining face processing (Wenger & Ingvalson, 2002; 2003; Richler et al., 2008; Cornes et al., 2011); thus, it could reveal evidence for violations of PI in these tasks.

The level of reliability for reporting violations of DS from marginal MSDA has been questioned. Mack, Richler, Gauthier and Palmeri (2011) demonstrated a reliance on the inferences about PS to be accurate before inferences about DS can be made. They suggested that violations of PS could be mischaracterised as violations of DS, leading to over-reporting of violations of DS in the marginal method. The point is a valid one. There are multiple methods to assess the three constructs of GRT and these different methods give varied accounts of the data. Menneer et al. (2010) modelled distributions for each violation scenario and ran simulations of the GRT analysis from three different methods of analysis. The proportion of times each violation was found by each model was reported. This showed that the marginal method was liberal to find violations of DS and the probit model was liberal to find violations of PI. Whist the MSDA method is advantageous in some respects, there are certain instances of uncertainty in interpretation of the results. Also, as outlined previously, the probit model may provide a better assessment of violations of PI. Therefore, computing and comparing multiple methods to assess all three constructs of GRT will hope to resolve discrepancies and uncertainty in the data. The MSDA measures used in the previous tasks, referred to as 'marginal measures' and the probit model will be used to estimate violations of the GRT constructs. Consistency will also be examined across the two measures as well as with previous evidence outlined in GRT tasks examining face processing.

Experiment 2.1

The aim of this project was to investigate three face processing tasks using quantitative methods that address formal definitions of configurality. Two of these tasks are analogues of composite face and Thatcher illusion tasks. The third is a task manipulating size, which is not related specifically to any standard face processing task but does belong to the family of generic manipulations made when comparing faces to probe faces. Cornes et al. (2011) found that there were multiple sources of the Thatcher illusion as demonstrated by differences in the pattern of violations of PS and DS across participants when an individual approach was used, but also in the simulations of the combinations of GRT inferences to provide evidence consistent with the Thatcher illusion. To allow GRT to explore the possibility for individual differences in evidence for configurality in these tasks, many trials per participant were required. Also, to allow the possible examination of consistency across time within participants, the study was repeated at three time points.

To analyse results at the level of the individual whilst trying to balance with the amount of time required to complete the study, the number of stimulus types in the design had been limited. The manipulations in each of the three tasks were: feature size of the eyes and mouth (matched to Wenger & Ingvalson, 2002; 2003), feature identity of the top and bottom half of the face (matched to Richler et al., 2008) and feature orientation of the eyes and mouth (matched to Cornes et al., 2011). For each of the three tasks, there were two versions: one with upright stimuli and one with inverted stimuli. One RI was used in the Feature Size Task and was set to match the longest RI used by Wenger and Ingvalson (2002, 2003). Only the aligned condition was used in the Feature Identity Task unlike Richler et al. (2008) and inverted faces served as the control rather than the addition of non-face objects. One hundred trials were needed for each stimulus type in order to compute the analysis. Therefore, 108 were displayed in order to allow for missed or slow responses that would later be omitted.

In the present study, a single group of participants were tested on three, two-alternative forced choice tasks. These tasks were similar to the three tasks previously investigated, characterised within a GRT framework. Choosing these tasks, and the stimuli manipulations used, was not to promote them as optimum, but instead to evaluate their suitability for demonstrating perceptual based configural processing at the behavioural level.

Aims

(1) To estimate, for each participant and in each condition, the magnitude of the between-feature, within-stimulus correlations using the probit methods in order to determine whether any evidence exists for the inference of within-stimulus configurality (violations of PI).

(2) To determine whether previous GRT findings could be replicated for these three tasks as these paradigms represent established manipulations to examine face processing.

(3) To see whether behavioural effects are replicated (e.g. inversion effects).

(4) To address whether the three tasks have different sources of configurality, with different patterns of GRT violations.

(5) To seek evidence for an orientation and status of the other feature interaction in the group analysis as evidence of a whole face processing advantage in the upright condition.

(6) To examine evidence for a selective marker of whole face processing in upright faces in the GRT violations.

(7) To explore the level of consistency across individuals by comparing the number of participants showing violations within each task.

Method

Participants

Seven postgraduate students at the University of Southampton volunteered to take part in the full study in return for payment. Four participants were female, five were White and three were Mixed Race. Participants had an age range of 22-25 years (M = 23.14, SD = 1.06). All participants had normal or corrected-to-normal vision. The researcher was one of these participants. Participant 6 was removed from analyses as the probit models were unable to converge on a stable solution of GRT parameter estimates. Thus, no results could be inferred about potential GRT violations for this participant's data.

Design

Three tasks were used in this experiment, with all observers performing all three tasks. Each task required participants to judge the status of two features (eyes and mouth, or top and bottom of the face) across two levels (either same versus different, or normal orientation versus odd orientation). Together, these two dimensions, each with two levels, created four stimulus conditions. These stimulus conditions were replicated in tasks requiring participants to judge feature size, identity and orientation. Each task was performed with upright and inverted faces. The set of tasks was then repeated three times by each participant, with a gap of approximately one month between repetitions.

In the Feature Size and Feature Identity Tasks, the eyes and the mouth (and top and bottom, respectively for the Feature Identity Task) were both judged for sameness (yes or no) in a successive matching task. In the Feature Orientation Task, participants judged whether eyes and mouths were the correct orientation relative to the face context (yes or no). For all tasks, judgments about the eyes (top) and the mouth (bottom) were made separately on each trial, but two responses were required on each trial. The order of task was the same across all participants: Feature Size, Feature Identity and then the Feature Orientation Task. The order of condition and which feature was responded to first was counterbalanced between participants, but remained the same across repeats (with random assignment of possible combinations). The response button was counterbalanced within participant and across repetitions.

Stimuli

Twenty-five (eleven male, fourteen female) faces from the NimStim face set (Tottenham et al., 2009) were selected to form a base stimulus set. The faces had no facial hair and blemishes were removed using Adobe Photoshop. Faces were manipulated to equate the positions of the pupil centres and the mouth across the images. Faces were placed within an oval annulus to mask hair and ears. Mean luminance and RMS contrast within the oval were then matched across all stimuli (Adams, Gray, Garner, & Graf, 2010).

In the Feature Size Task, 100 grey-scale stimuli were formed from the basic face set. Manipulated features were enlarged by 20% (see Wenger & Ingvalson, 2002, 2003, and Figure 2.2 for example stimuli). In the Feature Identity Task, composite faces were created from half faces, formed from the original stimulus set divided by a white line (3 pixel diameter) across the bridge of the nose. Only gender-consistent composite faces were formed. Some combinations were rejected due to the failure to make reasonable composites (e.g. bridge of the nose did not line up). After exclusions, 277 composite faces were created which were used to make 100 trial combinations. In the Feature Orientation Task, 100 stimuli were created. These consisted of 25 original grey-scale prepared faces, the same 25 faces with inverted eyes only, inverted mouths only or both features inverted. The eyes and mouths in the original stimulus set were manipulated as in Cornes et al. (2011). Also see Figure 2.2 for example stimuli.

Faces were presented centrally on the screen at a size of 3.70 cm by 5.00 cm at a viewing distance of 60 cm, creating approximate visual angles of 3.53° and 4.77°. For the Feature Identity Task, this visual angle was 3.53° by 4.10° (6.68° by 4.10° including the white divider line) as the top of the forehead was masked to remove the hairline.

In the Feature Size Task, a dot counting task was shown between study and probe faces to reduce the tendency to verbalise responses about size of features in the study face. This was not an issue for the Feature Identity Task as the feature components could not be verbalised in this way and the Feature Orientation Task was classification only (no study face). Dot stimuli in the feature size task were formed from white dots on a black background. The number of dots in the display ranged from one to eight and the size of each individual dot was approximately 0.40 cm diameter. The display of dots subtended visual angles of 6.90° by 8.30° or less in all cases.

The noise mask used throughout the tasks was created using the Gaussian monochromatic noise filter in Adobe Photoshop CS4 (see Figure 2.3). It appeared centrally on the screen with a luminance of 15.7 cd/m^2 comprised of white (36.20 cd/m^2) and black (0.11 cd/m^2) pixels) at a size of 7.00cm by 5.00 cm.

Apparatus and Materials

The tasks were built in Experiment Builder (Version 1.5.201). All stimuli were presented against a black background with a screen size of 36.50 cm by 27.50 cm, resolution of 1024 by 768 and refresh rate of 100 Hz. Responses were made via a mouse-button press. All text was presented in white. Prompts when responses were required consisted of ‘yes’ and ‘no’ being displayed on the side of the screen corresponding to the correct mouse button, along with a prompt for which question to respond to first. Testing sessions were run in dark room, and participants were seated at a distance of 60 cm from the screen with their head position maintained using a chin rest.

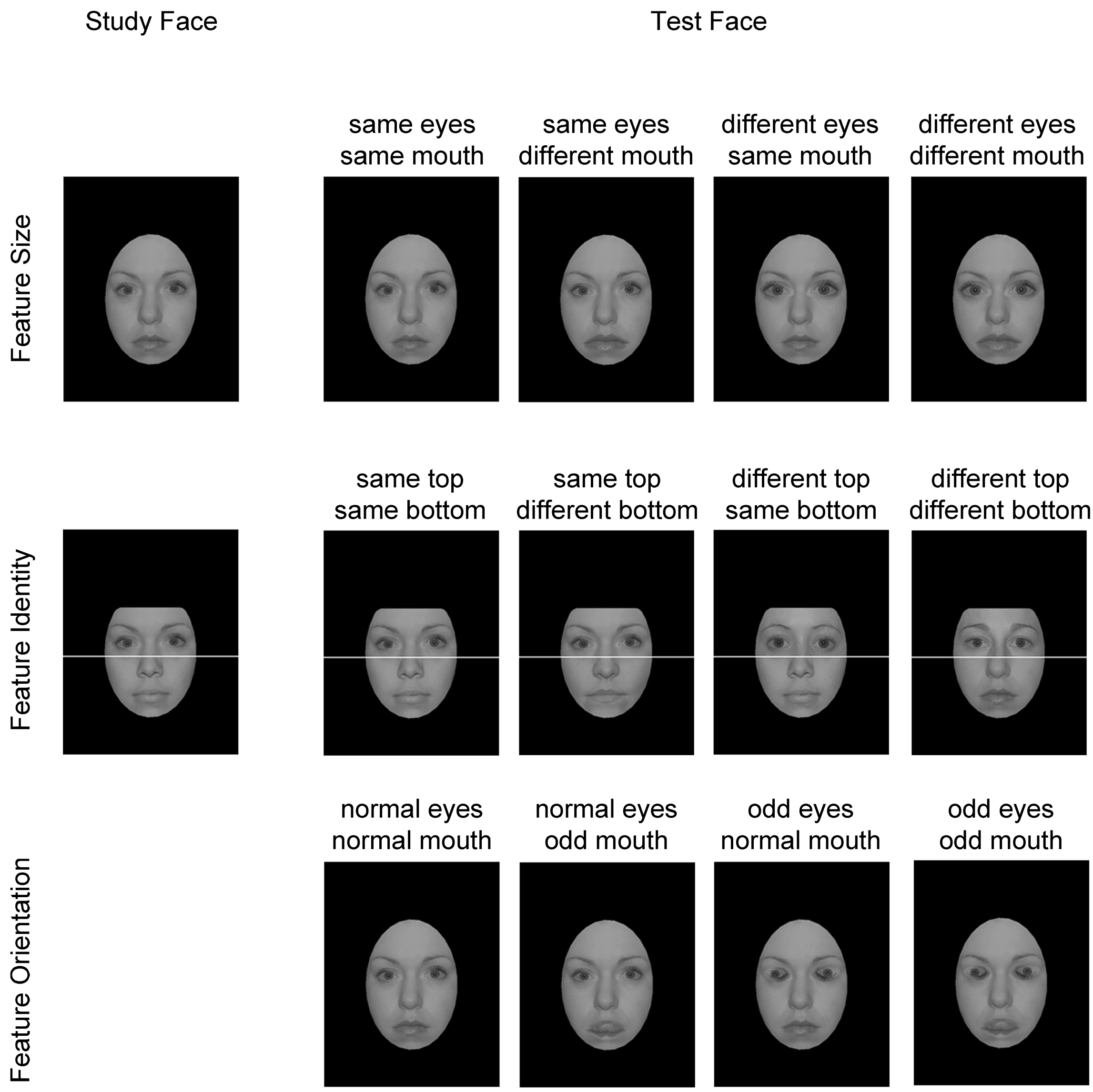


Figure 2.2. Examples of stimuli from each of the four conditions in the three tasks.

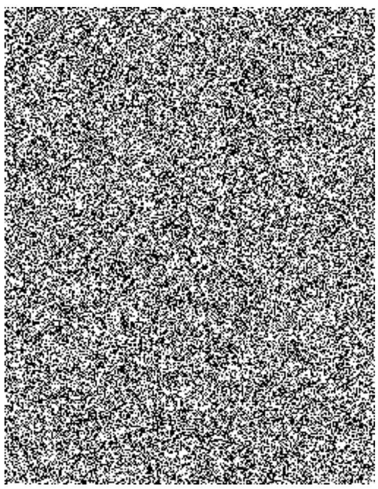


Figure 2.3. Example of mask stimulus used across all tasks.

Procedure

Participants completed three separate tasks. Three sessions (one for each task) were completed on successive days. Within each session, the upright and inverted conditions of the task were completed in blocks. Each orientation condition of a task contained eight blocks of 54 trials composed of 108 trials of each of the four trial types. The 25 possible stimuli in each trial type were shown at least 4 times. Ten practice trials (data were not recorded) were completed before the 432 experimental trials in each orientation condition, these also acted as adaptation trials to the dark room. Participants could take short self-paced breaks between blocks. The three-day experimental cycle was repeated after one month, then once again two months later to give a total of three repetitions of each task.

All trials were randomised within each session and began with a 500 ms fixation cross requiring participants to look at the centre of the screen. Trials ended with a 100 ms noise mask presented after both responses were made. No feedback was given. Only trials where the second response was made within three seconds of the first response were analysed (see Wenger & Ingvalson, 2002). The three tasks were designed to be similar to the procedures used in previous studies (see Experiment 1 of Wenger & Ingvalson, 2002, 2003; Richler et al., 2008; Cornes et al., 2011) while still providing similarities to each other.

Feature Size Task. Participants decided whether both the eyes and mouths of study and test faces were the same size (yes or no). Study faces were presented for 3000 ms. A 100 ms mask, and a dot counting task (200 ms display time, with associated response time) were presented between study and test faces. Participants responded either yes or no to a question about the number of dots that had appeared. Finally, there were an equal number of yes and no responses. Finally, the test face was displayed and remained visible until both responses were made. The dot counting task was used to ensure participants sub-vocalised the question rather than the appearance of the stimuli in the study face. The number of dots displayed and the number asked in the question were balanced to create the same number of yes and no answers across the task. The same number of yes and no responses were balanced across each condition to prevent patterns emerging with reference to responses on the main task. Accuracy at the counting task was also measured to ensure the distractor task was performed.

Feature Identity Task. Participants had to decide whether both the top and bottom halves of sequentially presented composite faces were the same (yes or no). Study faces were presented for 400 ms, followed by a 2000 ms mask and then the test faces. Test faces remained visible until both responses were made.

Feature Orientation Task. Participants decided if both the eyes and mouth were in the correct orientation relative to the face context (yes or no). If so, the terminology that the features were 'normal' is used. If not, then features are described as 'odd'. Faces were presented for 120 ms and were forward- and backward-masked with a 100 ms noise stimulus.

Note that when comparing the Feature Orientation to the Feature Size and Identity Tasks, 'same' trials are being mapped to 'normal' orientation and 'different' trials are being mapped to 'odd' orientation trials. The mapping might seem arbitrary; nevertheless, the reasoning is that finding differences between study and test faces is closer to identifying Thatcherised features than identifying 'normal' features. This is because 'odd' features in Thatcherised faces, when compared to a mental face norm, would prompt a 'different' response. In other words, the 'study' face in the Feature Orientation Task is the mental representation of a prototypical face stored in memory.

Results

Calculating Marginal Sensitivity and Bias

Only the trials in which a response was made to both features and the second response was made within three seconds of the first response were selected for further analysis (to match to procedure of Wenger & Ingvalson, 2002). For each experiment, a 4 x 4 matrix of responses was formed based on the four types of trial present in each experiment, and the four possible responses that could be made on each trial. Eighteen matrices were created for each participant in each of the three tasks, in each of the two conditions and at each of the three time points. If there was a zero for any cell in the matrix, then one was added to all values in the matrix to allow computation of MSDA measures. Analysis of latency to respond was not conducted as it was not pertinent to the research questions outlined.

The marginal hit rates and marginal false alarm rates were calculated from each of the participant matrices. These values were used to calculate the sensitivity (d') and bias (c) for one feature when the status of the other feature was held constant at a particular level. The Feature Size Task had four marginal conditions: response to the eyes when status of the mouth was 'same' (yes), response to the eyes when status of the mouth was 'different' (no), response to the mouth when status of the eyes were 'same' (yes) and response to the mouth when status of the eyes were 'different' (no). For the Feature Identity Task, the conditions were the same for the top and bottom halves of the face respectively. The Feature Orientation Task had four marginal conditions: response to the eyes when status of the mouth was 'normal' (yes), response to the eyes when status of the mouth was 'odd' (no), response to the mouth when status of the eyes were 'normal' (yes) and response to the mouth when status of the eyes were 'odd' (no). The '2AFC' paradigm was used, with the yes response (same or normal) as the hit rate (HR). For details of how the marginal signal detection results were calculated see Figure 2.4 and Equations 2-5. The model assumes each of the distributions is normally distributed and that subjects are responding optimally; however, the model can cope with some variation from the assumptions.

		Response			
		YY	YN	NY	NN
Stimulus	YY	HR (1) HR (3)	HR (1)	HR (3)	
	YN	HR (2) FA (3)	HR (2)	FA (3)	
	NY	FA (1) HR (4)	FA (1)	HR (4)	
	NN	FA (2) FA (4)	FA (2)	FA (4)	

HR = used to calculate hit rate
FA = used to calculate false alarm rate

(1) Marginal condition 1: A when B = yes
(2) Marginal condition 1: A when B = no
(3) Marginal condition 1: B when A = yes
(4) Marginal condition 1: B when A = no

A = eyes (top)
B = mouth (bottom)

Figure 2.4. Matrix and notation displaying the cells used to calculate the hit rate and false alarm rates for the marginal sensitivity and bias in each condition.

$$H = P(a_S\, b_i \mid A_S\, B_i) \qquad (2)$$

Marginal probability of correctly identifying the top element as being the same ('hit') when the bottom element is in state $_i$ (Same, Different)

$$F = P(a_S\, b_i \mid A_D\, B_i) \qquad (3)$$

Marginal probability of incorrectly identifying the top element as being the same ('false alarm') when the bottom element is in state $_i$ (Same, Different)

$$d' = \frac{1}{\sqrt{2}}\,[z(H) - z(F)] \qquad (4)$$

Where H is the hit rate, and F is the false alarm rate

$$d'\,(A, B_i) = \frac{1}{\sqrt{2}}\,\{z[P(a_S\, b_i \mid A_S\, B_i)] - z[P(a_S\, b_i \mid A_D\, B_i)]\}$$

Marginal sensitivity to the correct state of the top element when the bottom element is in state $_i$ (Same, Different)

$$c = -\frac{1}{2}\,[z(H) + z(F)] \qquad (5)$$

Where H is the hit rate, and F is the false alarm rate

$$c\,(A, B_i) = -\frac{1}{2}\,\{z[P(a_S\, b_i \mid A_S\, B_i)] + z[P(a_S\, b_i \mid A_D\, B_i)]\}$$

Marginal criterion for reporting the top element as the same when the bottom element is in state $_i$ (Same, Different)

Equation Notes. Lower case letters refer to responses, uppercase letters refer to stimuli. A refers to top elements (eyes, top half of face), B refers to bottom elements (mouth, bottom half of face). S as a subscript indicates 'same' status (or normal for the Feature Orientation Task), D as a subscript indicates 'different' status (or not normal for the Feature Orientation Task). All signal detection measures refer to the '2AFC' paradigm.

Consistency Across Time

The initial design of the study included separate time points for full data collection so that consistency of GRT evidence for configurality over time could be compared across the same participants. However, to evaluate consistency across time, there must be no changes in levels of performance across the time points. If there are differences in levels of performance over time then any inconsistency in sources for configurality in a task over time cannot be distinguished from differential performance due to learning or exposure effects of the stimuli. To evaluate whether there was a change in levels of performance, a series of six repeated measures ANOVAs (one for each task and orientation version) compared mean marginal d' for each participant across the three time points. Normality was checked using the Shapiro-Wilk test as samples across the conditions in each of the tests were less than 50. The statistic only reached significance in 1 of the 18 cases (p = 0.017); therefore, normality was not violated.

For the upright version of the Feature Size Task, the repeated measures ANOVA was significant (F(2,12) = 5.49, p = 0.020, η_p^2 = 0.48). Pairwise comparisons revealed no significant differences, but there was a significant linear trend between time one (M = 1.74, SE = 0.29) and time three (M = 2.08, SE = 0.35, F(1,6) = 7.61, p = 0.033, η_p^2 = 0.56). For the inverted version of the Feature Size Task the repeated measures ANOVA was significant (F(2,12) = 3.88, p = 0.050, η_p^2 = 0.39). Pairwise comparisons revealed a significant difference between d' at time one (M = 1.73, SE = 0.30) and time three (M = 2.05, SE = 0.37), p = 0.044. Comparisons with time two (M = 1.80, SE = 0.33) were not significant.

For the upright version of the Feature Identity Task, the repeated measures ANOVA was significant (F(2,12) = 7.85, p = 0.007, η_p^2 = 0.57). Pairwise comparisons revealed a significant difference between d' at time two (M = 1.29, SE = 0.19) and time three (M = 1.46, SE = 0.22), p = 0.027. Comparisons with time one (M = 1.19, SE = 0.18) were not significant. For the inverted version of the Feature Identity Task, the repeated measures ANOVA was non-significant (F(2,12) = 2.41, p = 0.132, η_p^2 = 0.28).

For the upright version of the Feature Orientation Task, the repeated measures ANOVA was significant (F(2,12) = 9.00, p = 0.004, η_p^2 = 0.60). Pairwise comparisons revealed significant differences between time one (M = 2.18, SE = 0.25) and times two (M = 2.33, SE = 0.26, p = 0.029) and three (M = 2.54, SE = 0.19, p = 0.039). For the inverted version of the Feature Orientation Task the repeated measures ANOVA was significant (F(2,12) = 7.77, p = 0.007, η_p^2 = 0.56). The pairwise comparisons revealed no significant differences, but there were significant linear trends between time one (M = 0.48, SE = 0.17) and time two (M = 0.66, SE = 0.15, F(1,6) = 9.10, p = 0.023, η_p^2 = 0.60) and between time one and time three (M = 0.91, SE = 0.16, F(1,6) = 7.38, p = 0.035, η_p^2 = 0.55).

As the performance appeared to improve over time in all tasks, with the exception of the inverted version of the Feature Identity Task, it was not appropriate to compare violations over time. Therefore, it was decided to combine trials over time points to create more robust data sets. This created a matrix of trials for each of the upright and inverted conditions of each of the three tasks, creating six matrices in total for each participant.

Task Performance

To check task performance of each of the participants, the individual d' and c results were examined. Firstly, marginal d' was examined for chance rate performance (d' of 0). The only condition in which marginal d' was at chance was the inverted version of the Feature Orientation Task (Figure 2.5). Previous studies measuring this same condition also show performance close to chance (Cornes et al., 2011); performance at chance was expected for this condition which demonstrates the Thatcher illusion (Thompson, 1980). From examination of the individual marginal d' results, it appears some participants are more sensitive than others in some tasks, but there do not appear to be consistent high and low performers across all tasks. The levels of sensitivity differ across tasks, with the Feature Identity Task showing the lowest levels of sensitivity across participants. There are also differences in marginal d' between orientations of a task, but the size of the difference varies by task. Sensitivity in the upright version of the Feature Orientation Task is higher than for the inverted version of the task across all participants.

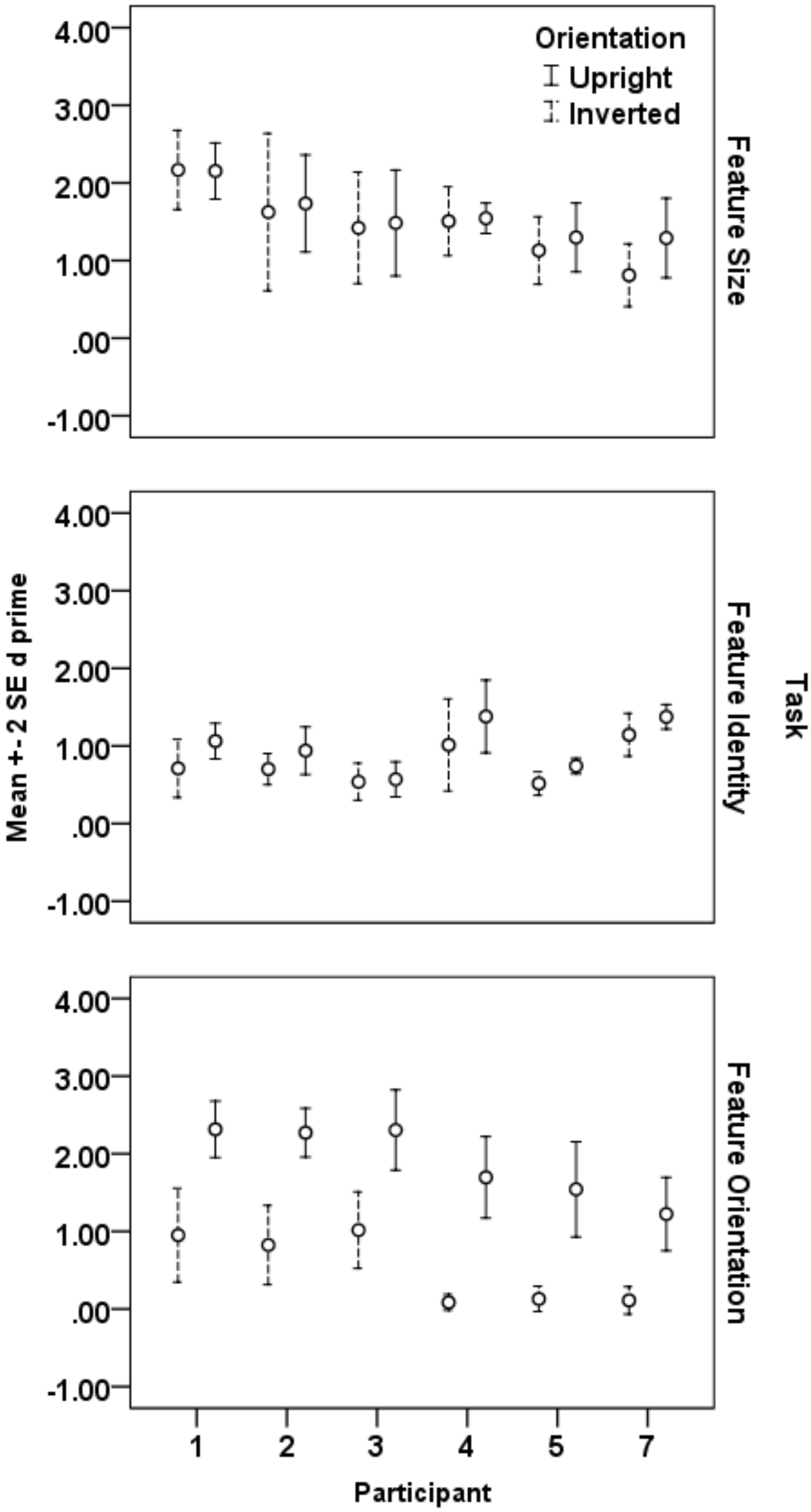


Figure 2.5. Error bar plots showing the marginal sensitivity (d' ± 2 SE) values for each participant in each condition. Plots are panelled by task (rows) and individual graphs are split by orientation.

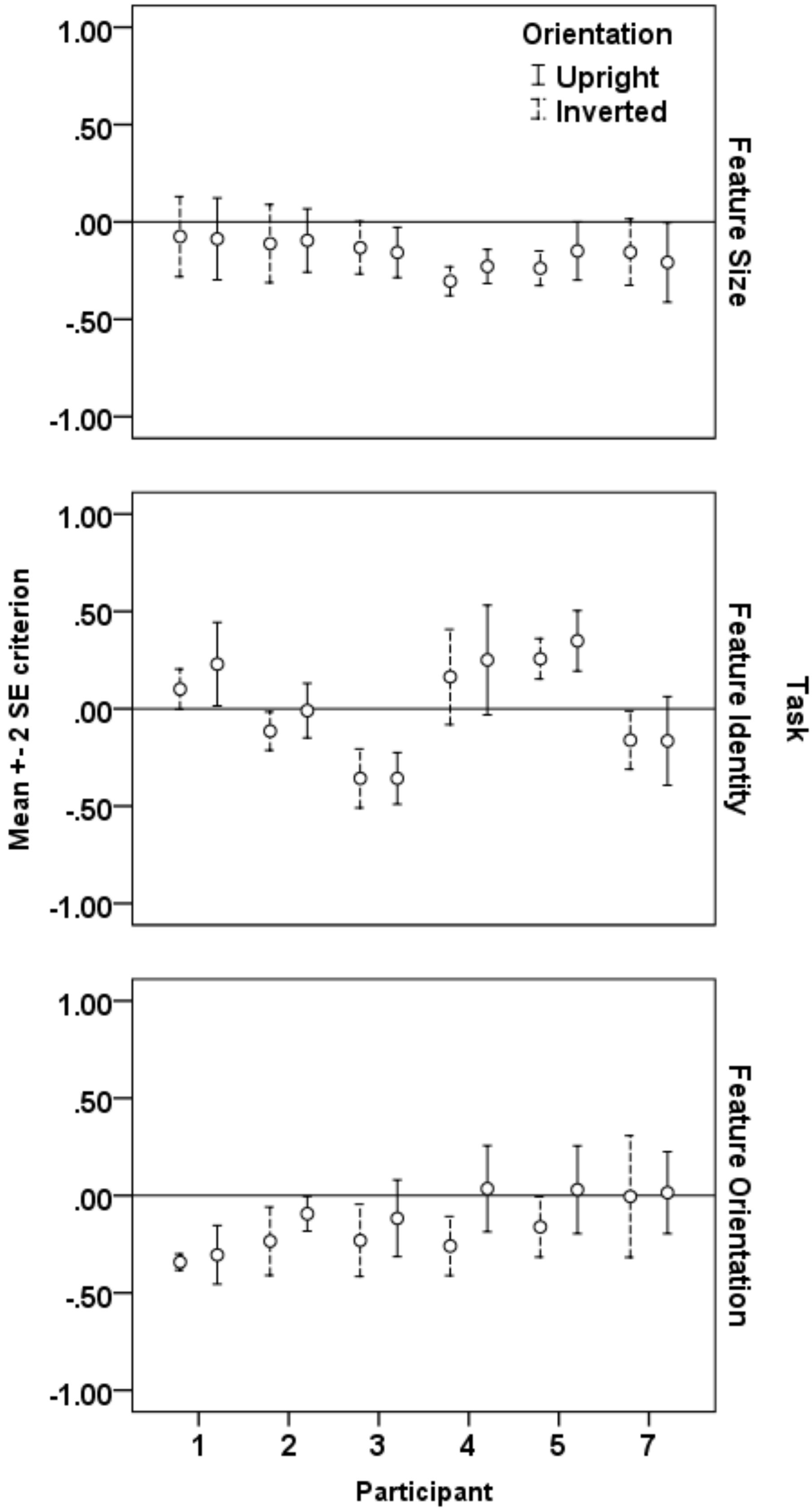


Figure 2.6. Error bar plots showing the marginal bias (criterion ± 2 SE) values for each participant in each condition. Plots are panelled by task (rows) and individual graphs are split by orientation.

For examining marginal bias, the further the c score is from zero (positive or negative), the larger the bias. Positive c values arise when the when false alarm rates are lower than miss rates (a conservative bias). Negative criterion values arise when the false alarm rate exceeds the miss rate (a liberal bias). Individual bias (criterion) results were examined in Figure 2.6. The Feature Size Task shows very low bias across participants and the values tend towards liberal bias. For the Feature Identity Task, there are higher levels of bias and more variation in bias between participants, with some participants showing liberal and some showing conservative bias. For the Feature Orientation Task, there is low bias or liberal bias demonstrated by participants. There is evidence for a conservative shift in bias between inverted and upright versions of the Feature Identity and Feature Orientation Tasks for some participants.

Group Analysis, MSDA and GRT Inferences

The task performance descriptive results reveal all participants were performing above chance level and were performing at similar levels in terms of sensitivity and bias in each of the tasks. These data were then analysed in two ways. First, estimates of marginal d' and c were analysed across all participants using analysis of variance (ANOVA). The analyses allowed examination of any differences in marginal d' and c, and whether any differences are consistent across the four stimulus conditions. Specifically evidence was sought of an interaction of face orientation and status of the other feature (same/different, normal/odd). Such an interaction would be consistent with evidence for whole face processing specific to upright but not inverted faces. Second, marginal and probit analyses were used to characterise potential dependencies between features. These analyses were conducted at the level of the individual observer. Both analyses were used to explore potential violations and to examine the level of agreement across the two methods. Following these analyses, and as a consequence of finding broad agreement across individuals in the patterns of violations, a third analysis was conducted. This third analysis sought to locate the sources of violations of PI found in the second analysis by examining the rho values produced by the probit analyses.

Group Analysis of Signal Detection Measures

Marginal sensitivity and bias in each task were analysed in two separate 2 (orientation: upright or inverted) x 2 (feature: eyes/top or mouth/bottom) x 2 (status of the other feature: same/normal orientation or different/odd orientation) factorial repeated measures ANOVAs.[3] Normality was checked using the Shapiro-Wilk test as samples across the conditions in each of the tests were less than 50. The statistic was not significant (*p*s> 0.05) in all but four conditions (out of a possible 36) across the six data sets and therefore normality was not violated.

Feature Size Task

Although performance at the counting task was not formally analysed as it was only used to prevent sub vocalisation, error rates only accounted for 2.80% of trials across the six participants. Trials for further analysis were not filtered by this response.

With respect to sensitivity, the main effects of feature and status of the other feature were significant ($F(1,5) = 8.95$, $MSE = 0.337$, $p = 0.030$; $F(1,5) = 16.36$, $MSE = 0.022$, $p = 0.010$). Sensitivity was higher to the eyes ($M = 1.76$, $SE = 0.28$) than the mouth ($M = 1.26$, $SE = 0.14$). Sensitivity was higher when the other feature was the same ($M = 1.60$, $SE = 0.21$), rather than different ($M = 1.43$, $SE = 0.21$). The main effect of orientation ($F(1,5) = 0.83$, $MSE = 0.288$, $p = 0.404$) failed to reach significance. All interactions were non-significant (all *F*s < 3.31, all *MSEs* > 0.005, all *p*s > 0.129). With respect to bias, there were no significant main effects or two-way interactions (all *F*s < 4.56, all *MSEs* > 0.003, all *p*s > 0.593). The three-way interaction of orientation, feature and status of the other feature was significant ($F(1,5) = 7.34$, $MSE = 0.002$, $p = 0.042$), showing a differential effect of status of the other feature across the eyes and mouth when upright, but not when inverted (see Figure 2.7).

[3] Mauchly's sphericity test could not be computed for any of these factors as df = 1 (i.e, if k = 2). When k = 2, sphericity is always met as there is only one variance of the difference between the levels. Therefore, sphericity cannot be violated for factors with two levels.

Feature Identity Task

With respect to sensitivity, the main effect of orientation was significant ($F(1,5) = 24.11$, $MSE = 0.029$, $p = 0.004$). Participants were more sensitive to upright ($M = 1.01$, $SE = 0.13$) than inverted faces ($M = 0.77$, $SE = 0.10$). No other main effects and, importantly, no interactions reached significance, (all Fs < 4.80, all $MSEs$ > 0.004, all ps > 0.08).

With respect to bias, the main effects of orientation, feature and status of the other feature were significant ($F(1,5) = 8.83$, $MSE = 0.06$, $p = 0.031$; $F(1,5) = 17.67$, $MSE = 0.020$, $p = 0.008$; $F(1,5) = 22.38$, $MSE = 0.019$, $p = 0.005$). Participants were more likely to respond 'same' in the inverted ($M = -0.02$, $SE = 0.09$) than upright ($M = 0.05$, $SE = 0.11$) condition, to the bottom part ($M = -0.07$, $SE = 0.12$) than to the top part ($M = 0.10$, $SE = 0.09$), and when status of the other feature was 'same' ($M = -0.08$, $SE = 0.09$) than 'different' ($M = 0.11$, $SE = 0.12$). No interactions reached significance (all Fs < 3.19, all $MSEs$ > 0.003, all ps > 0.134, see Figure 2.7).

Feature Orientation Task

With respect to sensitivity, the main effects of orientation, feature and status of the other feature were significant ($F(1,5) = 88.80$, $MSE = 0.255$, $p < 0.001$; $F(1,5) = 31.49$, $MSE = 0.254$, $p = 0.002$; $F(1,5) = 329.78$, $MSE = 0.002$, $p < 0.001$). Sensitivity was higher to upright ($M = 1.89$, $SE = 0.15$) than inverted faces ($M = 0.52$, $SE = 0.11$), to eyes ($M = 1.61$, $SE = 0.16$) than mouths ($M = 0.80$, $SE = 0.11$), and when status of the other feature was 'normal' ($M = 1.32$, $SE = 0.11$) than 'odd' ($M = 1.10$, $SE = 0.11$). The interaction between orientation and status of the other feature was significant ($F(1,5) = 35.50$, $MSE = 0.010$, $p = 0.002$). Sensitivity was significantly higher for upright faces ($F(1,5) = 107.72$, $MSE = 0.009$, $p < 0.001$) when the other feature was 'normal' ($M = 2.09$, $SE = 0.15$) than 'odd' ($M = 1.69$, $SE = 0.16$). In contrast there was no significant effect of status of the other feature for inverted faces ($F(1,5) = 4$, $MSE = 0.003$, $p = 0.10$). The three-way interaction was also significant ($F(1,5) = 14.36$, $MSE = 0.011$, $p = 0.013$, see Figure 2.7) showing a greater effect of status of the other feature for mouths than eyes when upright compared to inverted . All other interactions were non-significant (all Fs < 3.41, all $MSEs$ > 0.035, all ps > 0.124).

With respect to bias, the main effects of orientation, feature and status of the other feature were significant ($F(1,5) = 8.92$, $MSE = 0.024$, $p = 0.031$; $F(1,5) = 13.02$, *MSE* = 0.024, *p* = 0.015; *F*(1,5) = 12.18, *MSE* = 0.003, *p* = 0.017). Participants were less likely to respond 'normal' orientation in the upright (*M* = -0.07, *SE* = 0.06) than inverted condition (*M* = -0.20, *SE* = 0.07), to the mouth (*M* = -0.06, *SE* = 0.08) than the eyes (*M* = -0.21, *SE* = 0.05) and when the other feature was 'odd' (*M* = -0.11, *SE* = 0.07) than 'normal' (*M* = -0.16, *SE* = 0.06). No interactions reached significance (all *F*s < 6.55, all *MSEs* > 0.004, all *p*s > 0.05; Figure 2.7).

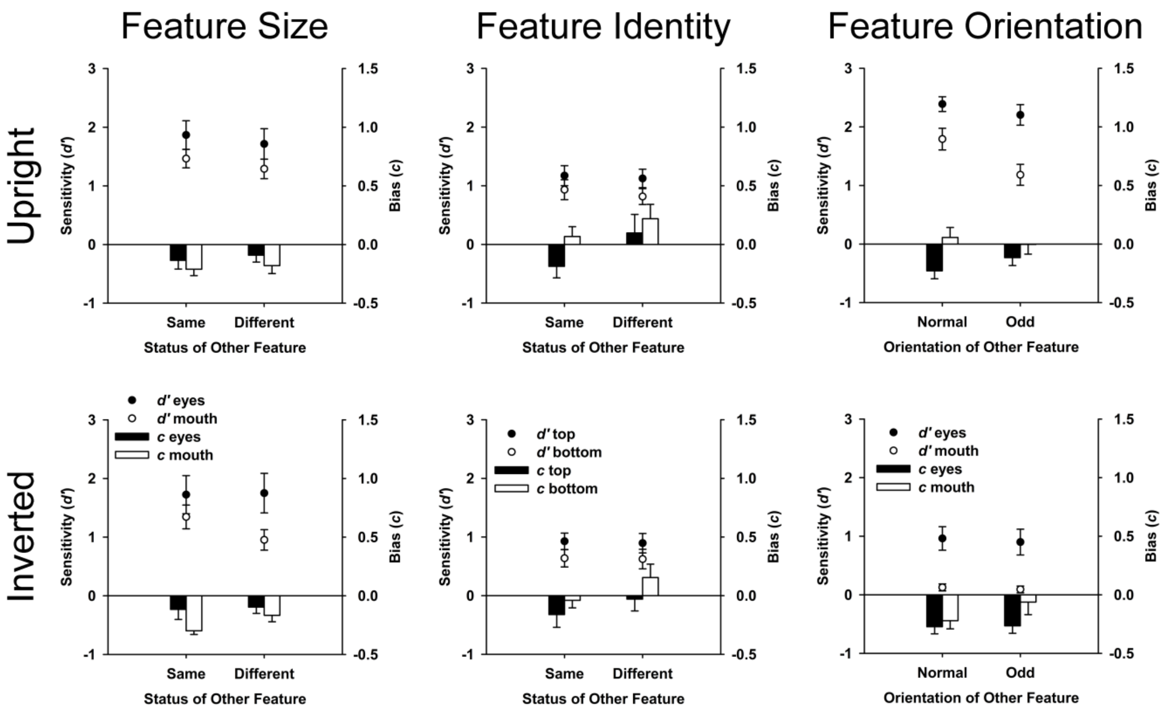


Figure 2.7. Plots of marginal sensitivity (d') and bias (c) with error bars representing standard error. Graphs paneled by task (Feature Size, Feature Identity, and Feature Orientation) and orientation condition (upright and inverted). Negative bias values indicate liberal bias to respond 'same' (Feature Size and Feature Identity Tasks) or 'normal' (Feature Orientation Task). Positive bias values indicate conservative bias, where participants are more likely to respond 'different' (Feature Size and Feature Identity Tasks) or 'odd' (Feature Orientation Task).

Group Analysis Summary

The analysis of group means demonstrated that sensitivity was influenced by the status of the other feature (Feature Size and Feature Orientation Tasks), orientation (Feature Identity and Feature Orientation Tasks) and feature (Feature Size and Feature Orientation Tasks). Crucially, the interaction between orientation and status of the other feature was significant in Feature Orientation Task only. In this case, 'normal' orientation in one feature enhanced sensitivity to the other feature when faces were upright, but not when inverted. The simplest account of this effect is that 'normal' orientation allows processing resources to be allocated to regions of potential 'inversion'. Seemingly, this allocation of resources can only be realised in upright faces.

With respect to bias, this was influenced by a common series of main effects and interactions in the Feature Identity and Feature Orientation Tasks. These data suggest a generalised effect of orientation, feature type and status of the other feature on decision-making in response to faces in these tasks.

MSDA and GRT Inferences

Marginal Analysis

Multidimensional signal detection analyses combine a set of non-parametric comparisons (Ashby & Townsend, 1986) and comparisons of parametric (typically Normal) measures of sensitivity and bias. They do so for one of the stimulus dimensions across levels of the other stimulus dimensions (Ashby & Townsend, 1986; Kadlec & Townsend, 1992b, 1992b), e.g. sensitivity to eyes across mouth-normal versus mouth-odd. Calculations of the measures of sensitivity and bias are performed in the same way as in one-dimensional signal detection theory (see MacMillan & Creelman, 2005). The results of these comparisons are combined (using the logic in, e.g. Kadlec & Townsend, 1992a, 1992b) to guide inferences regarding potential violations of PI, PS and DS.

Equality of marginal d' was calculated by creating a 95% confidence interval on the difference in d' to a feature across the two levels of the other feature (Equations 6 and 7). The calculation was corrected for multiple comparisons by using a z score of 2.243 as there were two comparisons (confidence intervals created for each of the two features across levels of the other feature). If the confidence interval contained zero then d' was equal across the two conditions (equality of d' is true). If the confidence interval did not contain zero then d' differed across the two conditions (equality of d' is false). Sensitivity to a feature differed depending on the level of the other feature.

$$\mathrm{var}(d') = \left(\frac{H(1-H)}{N_2[\phi(H)]^2}\right) + \left(\frac{F(1-F)}{N_1[\phi(F)]^2}\right) \quad (6)$$

Where N_2 and N_1 are the number of Signal (S_2) and noise (S_1) trials.
Where ϕ is the height of the normal density function at $z(p)$.

$$\text{Confidence Interval} = d'_d \pm z\left(\sqrt{\mathrm{var}(d'_1) + \mathrm{var}(d'_2)}\right) \quad (7)$$

Where d'_d is the difference between d' 1 and 2, $\mathrm{var}(d'_1)$ is the variance of d' 1, and $\mathrm{var}(d'_2)$ is the variance of d' 2.

Equality of marginal c was calculated in the same way, by creating a 95% confidence interval (with correction for two comparisons) on the difference in c to a feature across the two levels of the other feature (Equations 8 and 9). If the confidence interval contained zero then bias was equal across the two conditions (equality of c is true). If the confidence interval did not contain zero then bias differed across the two conditions (equality of c is false). Values of c and d' provide evidence for inferences of DS and PS respectively.

$$\mathrm{var}(c) = \frac{1}{4}\left(\left(\frac{H(1-H)}{N_2[\phi(H)]^2}\right) + \left(\frac{F(1-F)}{N_1[\phi(F)]^2}\right)\right) \quad (8)$$

Where N_2 and N_1 are the number of Signal (S_2) and noise (S_1) trials.
Where ϕ is the height of the normal density function at $z(p)$.

$$\text{Confidence Interval} = c_d \pm z\left(\sqrt{\mathrm{var}(c_1) + \mathrm{var}(c_2)}\right) \quad (9)$$

Where c_d is the difference between criterion 1 and 2, $\mathrm{var}(c_1)$ is the variance of criterion 1, and $\mathrm{var}(c_2)$ is the variance of criterion 2.

For i = 1,2: (10)

$$P(R_{xiy1}|X_iY_1) + P(R_{xiy2}|X_iY_1) = P(R_{xiy1}|X_iY_2) + P(R_{xiy2}|X_iY_2)$$

For j = 1,2:

$$P(R_{x1yj}|X_1Y_j) + P(R_{x2yj}|X_1Y_j) = P(R_{x1yj}|X_2Y_j) + P(R_{x2yj}|X_2Y_j)$$

Marginal response invariance: Non-parametric equality testing for the presence of marginal response invariance, pertinent to inferences regarding PS and DS.

A non-parametric test of marginal response invariance is also used, in conjunction with the marginal measures of sensitivity and bias, to determine whether DS and PS hold, using equalities calculated for each of the two dimensions (Equation 10). These equalities check whether the probability of responding 1 or 2 (i.e. 'same' or 'different', 'normal' or 'odd') in the *y*-dimension is the same when Y = 1 as it is when Y = 2, and similarly check with the probability of responding 1 or 2 in the *x*-dimension is the same regardless of whether X = 1 or X = 2. If variance is found then marginal response invariance is false. If no variance is found, then marginal response invariance is true. If these equalities are satisfied (true) in the data, then tests of equality of d' and c can be used to determine if PS, DS, or both are violated. If the marginal response invariance equalities are not satisfied, then inferences regarding PS and DS become potentially problematic.

Table 2.1

Logic Relating Marginal Measures of Performance to Inferences Regarding PS and DS

Evidence			Inference	
MRI?	Equal marginal d'	Equal marginal c	PS	DS
T	T	T	T	T
T	T	F	T	F
T	F	T	F	T
T	F	F	F	F
F	T	T	T	?
F	T	F	T	F
F	F	T	F	?
F	F	F	F	?

Note. Marginal response invariance represented by MRI. T for the evidence indicates the equality in question held, T for the inference indicates no violation; F for the evidence indicates the equality in question did not hold, F for the evidence indicates a violation; ? indicates an uncertain inference (Cornes et al., 2011).

Inferences about PS and DS were made based on evidence from these three equalities (Table 2.1). Different combinations of evidence lead to different inferences regarding PS and DS. In some circumstances, there is uncertainty in inferences regarding DS and these will not be counted as violations. This uncertainty arises because there is a reliance on the inferences about PS to be accurate before inferences about DS can be made. For DS and PS there are two locations of measurement, in the horizontal, and in the vertical dimensions of the multivariate probability distributions (see Figure 2.8). A horizontal violation is for Feature B across levels of Feature A, and a vertical violation would be observed for Feature A across levels of Feature B.

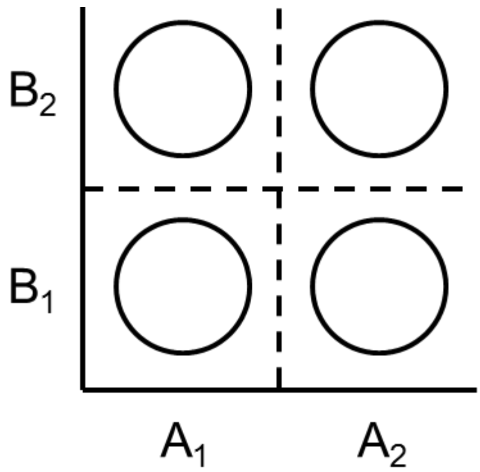


Figure 2.8. Diagram showing the multivariate probability distribution for each of the three tasks. For the Feature Size Task, Feature A is the eyes, and Feature B is the mouth; level 1 is 'same and level 2 is 'different'. For the Feature Identity Task, Feature A is the top half of the face and Feature B is the bottom half of the face; level 1 is 'same' and level 2 is 'different'. For the Feature Orientation Task, Feature A is the eyes and Feature B is the mouth; level 1 is 'normal orientation' and level 2 is 'inverted orientation'.

The potential correlation between dimensions is not addressed by one-dimensional SDT as this is an issue that only exists within multidimensionality. Therefore, inferences regarding potential violations of PI are assessed indirectly using a non-parametric test of sampling independence (Ashby & Townsend, 1986; Kadlec & Townsend, 1992a, 1992b) (Equation 11). If DS and PI hold, then the probability of responding X = 1 and Y = 1, for a given stimulus type (X_iY_j), is the joint probability of responding X = 1 and of responding Y = 1. If this equality is not satisfied (the sides of the equation are not equal) in the data for each stimulus type, then this provides evidence for a violation of PI, contingent on DS holding. Only one violation of the possible four is considered necessary to draw a conclusion about PI for the task as the probability equation is based on very robust data.

$$(R_{x1y1}|X_iY_j) = [P(R_{x1y1}|X_iY_j) + P(R_{x1y2}|X_iY_j)]\,[P(R_{x1y1}|X_iY_j) + P(R_{x2y1}|X_iY_j)] \tag{11}$$

Sampling independence: Non-parametric equality testing for the presence of sampling independence when the top element is the same and the bottom element is in state $_i$ (Same, Different), pertinent to inferences regarding PI.

These calculations differ from those published previously (e.g. Ashby & Townsend, 1986; Kadlec, 1995, 1999; Wenger & Ingvalson, 2002) and some micro analyses (conditional measures, Wenger & Ingvalson, 2002) are omitted as they are often non-normal. The analysis used in previous studies examining GRT principles in the face processing have based analysis on MSDA (Kadlec & Townsend, 1992a). This is referred to as 'marginal analysis' in this chapter due to the marginal measures used to make inferences.

Marginal Results

The individual participant results are presented in Table 2.2. The marginal analyses revealed three main findings: (1) no violations of PI in any task (other than for one participant in the inverted condition of the Feature Size Task); (2) frequent violations of PS in the upright condition of the Feature Orientation Task with modest numbers of violations in upright and inverted conditions of the Feature Size Task; and (3) frequent violations of DS, especially for the Feature Identity Task. From the marginal analysis, there does not appear to be great consistency across the tasks. There is some evidence for consistency of violations across participants, particularly for the upright Feature Orientation Task and both conditions of the Feature Identity Task.

Decisional effects identified may be overstated in some cases (criticisms over reporting violations of DS and PS, Mack et al., 2011), but there was not a single violation of PI found by the marginal analysis. The results support findings from previous studies (Wenger & Ingvalson, 2002, 2003; Richler et al., 2008; Cornes et al., 2011) that there are decisional sources for the effect of configurality. Also, there is no evidence for violations within a single stimulus (violations of PI) which would provide the strongest evidence for holism as defined by the holistic perception hypothesis (Farah, Wilson, Drain, & Tanaka, 1998).

Table 2.2

Summary of Marginal Violations for each Participant in each Task

			Marginal Inferences							
	Orientation	Participant	PI11	PI21	PI12	PI22	PSh	PSv	DSh	DSv
Feature Size Task	Upright	1	0	0	0	0	0	0	0	0
		2	0	0	0	0	1	0	?	0
		3	0	0	0	0	0	0	0	0
		4	0	0	0	0	0	1	0	?
		5	0	0	0	0	1	1	?	?
		7	0	0	0	0	0	1	0	?
	Inverted	1	0	0	0	0	0	0	0	0
		2	0	0	0	0	1	0	?	0
		3	0	0	0	0	0	0	0	0
		4	0	0	0	0	1	0	0	0
		5	0	0	0	0	0	0	1	0
		7	0	1	0	0	1	1	?	?
Feature Identity Task	Upright	1	0	0	0	0	1	0	?	1
		2	0	0	0	0	0	0	0	1
		3	0	0	0	0	0	0	0	0
		4	0	0	0	0	1	0	?	1
		5	0	0	0	0	0	0	1	1
		7	0	0	0	0	0	0	1	1
	Inverted	1	0	0	0	0	0	0	1	0
		2	0	0	0	0	0	0	1	0
		3	0	0	0	0	0	0	0	1
		4	0	0	0	0	0	0	1	0
		5	0	0	0	0	0	0	1	1
		7	0	0	0	0	0	0	1	1
Feature Orientation Task	Upright	1	0	0	0	0	1	1	?	?
		2	0	0	0	0	1	0	?	1
		3	0	0	0	0	1	0	?	0
		4	0	0	0	0	1	0	?	0
		5	0	0	0	0	1	0	?	0
		7	0	0	0	0	1	0	?	0
	Inverted	1	0	0	0	0	0	0	0	0
		2	0	0	0	0	0	0	1	0
		3	0	0	0	0	0	0	1	?
		4	0	0	0	0	1	0	?	0
		5	0	0	0	0	0	0	1	0
		7	0	0	0	0	0	0	0	0

Note. Violations of equalities and measures are noted as '1', the measure holding is noted as '0', '?' indicates an uncertain inference. For the numbers after the measures of PI, these refer to the particular distributions of which there are four. For the letters after the measures, 'h' denotes the measure in the horizontal (for the mouth across levels of the eyes or for the bottom across levels of the top for the Feature Identity Task), 'v' denoted the measure in the vertical (for the eyes across levels of the mouth or for the top across levels of the bottom for the Feature Identity Task). Violations of marginal inferences are highlighted in grey.

Probit Models

The marginal analyses were augmented with direct estimates of parameters of the underlying normal distributions and the decision bounds. The critical innovation in the present study is the use of probit regression (DeCarlo, 2003) to estimate these quantities. In previous research, probit models have been shown to be more sensitive than marginal analyses to the presence of within-stimulus correlations (Menneer, et al., 2010, 2012; Menneer et al., 2009).

The probit model approach used in this chapter uses a technique developed by DeCarlo (2003) in which probit models are used to determine the signal detection measures in multidimensional space using *z*-scores for response rates. DeCarlo's (2003) method was used to determine a two-dimensional configuration in which three distributions exist. Here, the modelling technique developed by Menneer et al. (2010) is applied to estimating the GRT parameters for a two-dimensional configuration with four distributions. This approach also differs from the DeCarlo (2003) approach as DS was not previously employed.

Probit models were implemented using two structures. In the first, each probit model was based on a single distribution (i.e. one for each stimulus), in order to estimate the criteria (cs) and the bivariate correlations for each distribution. In the second structure, each probit model was implemented over two neighbouring distributions in order to directly estimate d'. Both types of model were of the form given in Equation 11. The outcome y depends on the value of $y*$ relative to a criterion, c (Equation 12).

$$y^* = \beta x + \mu \tag{11}$$

$y*$ is the latent (dependent) variable; β is the regressor set for *x*, providing the *z*-score for the proportion of '1' responses, used to calculate d'; *x* is the explanatory variable (correct response); and μ is the residual distribution, providing bivariate correlations (ρ).

$$y = \begin{Bmatrix} 0\ if\ y* < c \\ 1\ if\ y* \geq c \end{Bmatrix} \tag{12}$$

c is the decision bound, estimated in two models, each with the same pair of distributions

In the first structure (one distribution per model), a linear model with a probit link function was implemented for each distribution, in each dimension. The data for each model was restricted to the response data for the given distribution. For each distribution, two models were implemented, one in the x-dimension and one in the y-dimension, giving eight models in total. Each criterion (c) was estimated across two models, one for each of the distributions either side of the criterion. The bivariate correlation for each distribution was estimated from residuals (μ) across two models, one for each dimension within the given distribution.

In the second structure (two distributions per model), a model was implemented for the two distributions in each level of each dimension. For example, for dimension y at level 1, there are two distributions: one at $x = 1$ and one at $x = 2$. By including both distributions in the model, the distance between the distribution means (d') can be estimated directly from β. In this way, marginal signal detection parameters can be estimated in a two-dimensional case in the same way as specified in DeCarlo (1998) for the one-dimensional case.

In both structures, the criteria were estimated separately for each level within a dimension; hence, DS was not enforced (unlike DeCarlo, 2003). Criteria, d's and correlations were estimated separately to avoid under-identification of the models.

Probit Results

The individual participant results from the probit models are presented in Table 2.3. There were three key findings: (1) The Feature Identity Task leads to frequent violations of DS in upright and inverted conditions, with violations of DS in the other tasks largely restricted to inverted conditions; (2) violations of PS are most commonly found in the upright condition of the Feature Orientation Task; and (3) violations of PI are reliably present in all tasks and conditions, although they were somewhat less common in the Feature Identity Task.

Table 2.3

Summary of Probit Violations for each Participant in each Task

			Probit Inferences							
	Orientation	Participant	PI11	PI21	PI12	PI22	PSh	PSv	DSh	DSv
Feature Size Task	Upright	1	0	1	1	0	0	0	0	0
		2	1	1	1	0	0	0	0	0
		3	1	1	0	1	0	0	0	0
		4	1	0	1	1	0	0	0	0
		5	0	0	0	0	0	0	0	0
		7	0	1	1	0	0	0	0	0
	Inverted	1	1	1	1	0	0	0	0	0
		2	1	0	0	1	0	0	0	0
		3	0	1	0	0	0	0	0	0
		4	1	0	0	0	0	0	0	0
		5	0	0	1	0	0	0	1	0
		7	0	1	1	0	1	0	-	1
Feature Identity Task	Upright	1	1	0	0	0	1	0	-	1
		2	0	1	0	0	0	0	0	1
		3	0	1	1	1	0	0	0	0
		4	1	0	0	0	1	0	-	1
		5	0	0	0	0	0	0	1	1
		7	0	0	0	0	0	0	0	0
	Inverted	1	1	0	1	0	0	0	1	0
		2	0	1	0	0	1	0	-	0
		3	0	0	0	0	0	0	0	1
		4	1	1	0	0	0	0	1	0
		5	0	0	0	0	0	0	1	1
		7	0	0	0	0	0	0	0	1
Feature Orientation Task	Upright	1	0	0	1	0	0	0	0	0
		2	1	0	0	1	1	0	-	0
		3	0	1	0	1	1	0	-	0
		4	1	0	1	0	1	0	-	0
		5	0	1	0	0	1	0	-	0
		7	1	1	1	0	1	0	-	0
	Inverted	1	0	0	0	0	0	0	0	0
		2	0	0	0	1	0	0	1	0
		3	1	1	0	1	0	0	1	0
		4	0	0	0	0	1	0	-	0
		5	0	0	0	1	0	0	1	0
		7	0	0	0	1	0	0	0	0

Note. Violations of equalities and measures are noted as '1', the measure holding is noted as '0', '-' indicates the model could not converge for this measure. For the numbers after the measures of PI, these refer to the particular distributions of which there are four. For the letters after the measures, 'h' denotes the measure in the horizontal (for the mouth across levels of the eyes or for the bottom across levels of the top for the Feature Identity Task), 'v' denoted the measure in the vertical (for the eyes across levels of the mouth or for the top across levels of the bottom for the Feature Identity Task). Violations of probit inferences are highlighted in grey.

GRT Results Summary

A summary of violations from the marginal and probit analyses are presented in Figure 2.9. There can be up to four instances of violations of PI for each participant. If at least one is shown then a violation of PI was recorded for that participant in this figure. For violations of PS and DS there are two possible sources for each participant. Again, if at least one violation is shown, then a violation is recorded for that participant in the figure. The analyses motivated by GRT provide further, and different, insights from those available from group means. The marginal and probit analyses converge to suggest three findings: (1) violations of PS to upright but not inverted faces in the Feature Orientation Task, (2) frequent violations of DS in upright and inverted conditions of the Feature Identity Task and (3) some violations of DS in the inverted conditions of Feature Size and Orientation Tasks found across analyses.

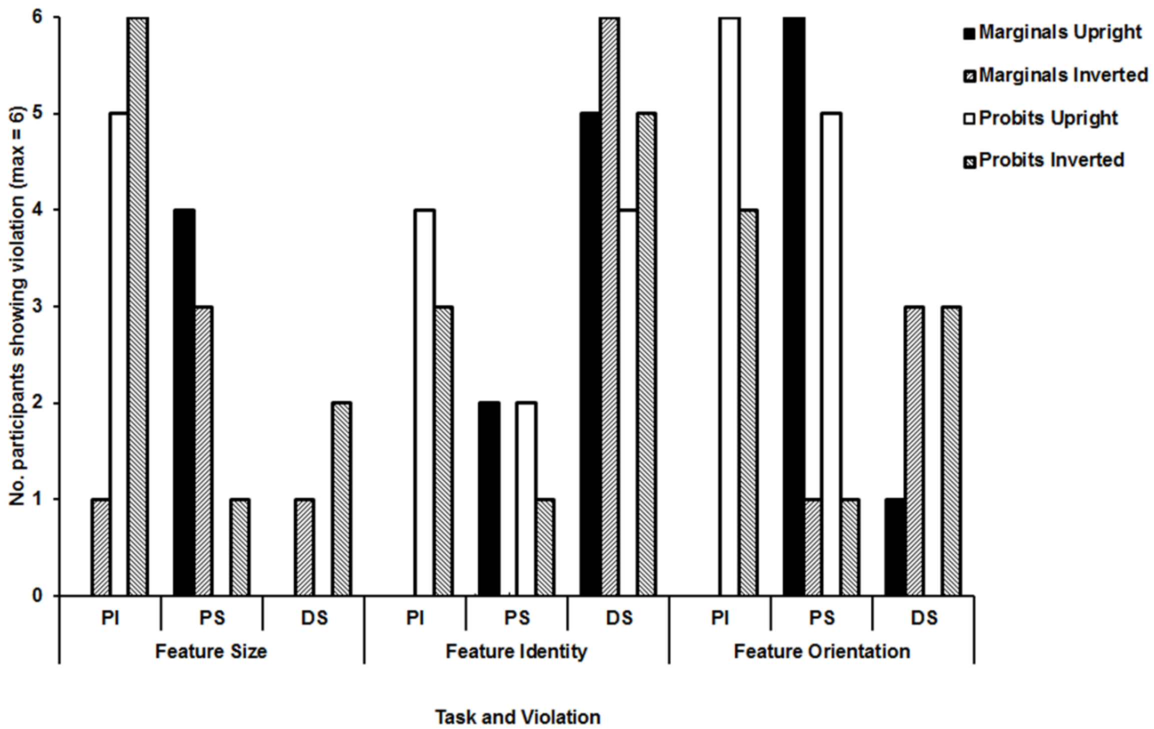


Figure 2.9. Summary of significant violations in marginal and probit analyses for each of the tasks and orientation conditions.

These results confirm an effect of orientation on sensitivity in the Feature Orientation Task that is not present in either of the other tasks. They also confirm that the Feature Identity Task is influenced by shifts in response criterion across conditions. Additionally, there is evidence that face inversion is associated with a more general tendency to shift criterion across conditions. This latter finding is consistent with face inversion creating a situation where decisions to faces are subject to problem solving strategies.

Evidence was found for violations of PI almost exclusively in the probit analyses. Frequent violations were found in the Feature Size and Orientation tasks, in both upright and inverted conditions, with a reduced number in the Feature Identity Task. The marginal analyses suggested only a single violation of PI. The discrepancy is a cause for concern and might reflect differences in the tendency to make Type I and Type II errors rather than differences in sensitivity.

Confidence that the difference between probit and marginal analyses reflects sensitivity to violations rather than a tendency to make Type I errors in the case of the probit analysis is strengthened by simulations that have been performed by Tamaryn Menneer in order to compare the two approaches (Menneer et al., 2012; 2009; Mestry, Menneer et al., 2012). In Mestry, Menneer et al. (2012), simulated distributions containing known violations of PI, PS and DS of various magnitudes, pair-wise combinations of violations, and combinations of all three. A total of 1,000 confusion matrices were generated for each violation condition. From each distribution, 250 points were sampled, each of which represents a participant response to a given trial of a given stimulus type. This number is comparable to the 300 trials that were used for each stimulus type in the current experiment. Simulations were created assuming a true d' of 1 or 2 in order to approximate the values found in the current experimental data. Using these data, the ability of the probit and marginal models to detect these known violations was tested. The results of these simulations are presented in Appendix B.

For the marginal analyses, results show Type II errors for violations of PI. For the probit analyses, results show Type I errors for violations of PI but only when there is a violation of DS with a continuous decision bound (i.e. when the change in the decision bound from one level to the next is continuous, rather than a step function). When this type of violation of DS occurs, correlations in the response confusion matrix can appear as violations of PI with the same sign and similar magnitude in all distributions. Table 2.4 contains the mean correlation estimates from the current experimental data. These correlations are not of the same sign within each set of distributions. Therefore, it is unlikely that the violations of PI reported in the probit analysis are Type I errors.

In conclusion, the probit analyses are more sensitive to violations of PI compared to the marginal analyses. Somewhat troublingly for the vernacular conception of configural processing, the violations of PI reported are found as frequently in inverted faces as in upright faces. Although these results do support the fact that humans do compute between-feature relationships in faces, they do not support the fact that these computations are made for upright but not inverted faces. However, there remains one further possibility that violations of PI are orientation specific. As is standard, the existence of a violation of PI was reported if at least one (out of four) of the bivariate distributions (i.e. stimulus types) exhibits a significant correlation of its underlying dimensions. It is possible that the orientation specificity of violations of PI exists in differences in the bivariate distributions that show correlations across dimensions.

Correlation Analysis

The mean correlations for bivariate distributions are presented in Table 2.4. The relative positions of the bivariate distributions and decision criteria in stimulus space for the averaged data are presented in Table 2.5 for both marginal and probit analyses, and graphically for probit analyses in Figure 2.10. In these analyses, dimension X is the eye/top feature of the face and dimension Y is the mouth/bottom feature of the face; level 1 is same/normal orientation and level 2 is different/odd orientation. When these correlation values were subjected to ANOVA, the interaction between orientation and distribution was significant in both the Feature Identity Task ($F(3,15) = 5.14$, $MSE = 0.006$, $p = 0.012$) and Feature Orientation Task ($F(3,15) = 14.03$, $MSE = 0.039$, $p < 0.001$), but not the Feature Size Task ($F(3,15) = 1.16$, $MSE = 0.018$, $p = 0.357$).

Table 2.4

Mean Correlation for each Distribution in each Task Condition with 95% Confidence Intervals

Task	Orientation	Distribution	*M*	*SE*	Lower CI	Upper CI	Different from zero
Feature Size	Upright	X_1Y_1	0.34	0.10	0.08	0.60	*
		X_2Y_1	-0.40	0.08	-0.59	-0.21	*
		X_1Y_2	-0.37	0.08	-0.58	-0.16	*
		X_2Y_2	0.08	0.11	-0.20	0.36	
	Inverted	X_1Y_1	0.25	0.13	-0.09	0.58	
		X_2Y_1	-0.40	0.13	-0.71	-0.09	*
		X_1Y_2	-0.26	0.03	-0.33	-0.18	*
		X_2Y_2	0.08	0.09	-0.15	0.31	
Feature Identity	Upright	X_1Y_1	0.19	0.08	0.00	0.38	
		X_2Y_1	-0.11	0.12	-0.40	0.18	
		X_1Y_2	-0.09	0.07	-0.28	0.10	
		X_2Y_2	-0.03	0.09	-0.26	0.20	
	Inverted	X_1Y_1	0.12	0.07	-0.05	0.29	
		X_2Y_1	-0.02	0.10	-0.26	0.22	
		X_1Y_2	0.07	0.09	-0.15	0.28	
		X_2Y_2	-0.02	0.06	-0.16	0.13	
Feature Orientation	Upright	X_1Y_1	0.46	0.10	0.20	0.72	*
		X_2Y_1	-0.41	0.09	-0.63	-0.19	*
		X_1Y_2	-0.43	0.11	-0.70	-0.17	*
		X_2Y_2	-0.07	0.11	-0.35	0.21	
	Inverted	X_1Y_1	-0.04	0.09	-0.26	0.17	
		X_2Y_1	-0.10	0.05	-0.23	0.03	
		X_1Y_2	-0.03	0.06	-0.18	0.13	
		X_2Y_2	-0.24	0.06	-0.39	-0.08	*

Note. * indicates a significant difference at $p < 0.05$. Grey highlights indicate the distributions in which there is an inversion effect as the mean rho from the distribution in one orientation does not fall into the confidence interval of the rho value in the other orientation distribution.

Table 2.5

Sensitivity and Bias values (and standard errors) for the Marginal and Probit Analyses

			Sensitivity (d')				Bias (c)			
			X_1Y_1 to X_2Y_1	X_1Y_1 to X_1Y_2	X_2Y_1 to X_2Y_2	X_1Y_2 to X_2Y_2	X_1Y_1 to X_2Y_1	X_1Y_1 to X_1Y_2	X_2Y_1 to X_2Y_2	X_1Y_2 to X_2Y_2
Marginals	Upright	FS	1.87	1.46	1.29	1.71	-0.13	-0.21	-0.18	-0.09
			(0.23)	(0.08)	(0.18)	(0.15)	(0.08)	(0.04)	(0.05)	(0.06)
		FI	1.17	0.93	0.82	1.12	-0.19	0.07	0.22	0.10
			(0.14)	(0.18)	(0.17)	(0.14)	(0.07)	(0.07)	(0.11)	(0.15)
		FO	2.39	1.79	1.18	2.20	-0.23	0.06	0.00	-0.12
			(0.29)	(0.30)	(0.22)	(0.29)	(0.06)	(0.09)	(0.09)	(0.06)
	Inverted	FS	1.72	1.34	0.95	1.75	-0.12	-0.30	-0.17	-0.10
			(0.27)	(0.14)	(0.16)	(0.27)	(0.10)	(0.05)	(0.07)	(0.06)
		FI	0.93	0.64	0.62	0.89	-0.16	-0.04	0.15	-0.03
			(0.11)	(0.16)	(0.19)	(0.14)	(0.10)	(0.05)	(0.12)	(0.08)
		FO	0.96	0.13	0.09	0.90	-0.27	-0.22	-0.06	-0.27
			(0.24)	(0.29)	(0.27)	(0.23)	(0.03)	(0.05)	(0.07)	(0.03)
Probits	Upright	FS	1.63	1.41	1.34	1.57	0.08	0.11	0.10	0.06
			(0.11)	(0.07)	(0.08)	(0.13)	(0.04)	(0.02)	(0.04)	(0.03)
		FI	1.31	1.16	1.07	1.28	0.13	-0.05	-0.18	-0.06
			(0.08)	(0.08)	(0.07)	(0.09)	(0.08)	(0.07)	(0.11)	(0.12)
		FO	1.80	1.64	1.32	1.79	0.07	-0.01	0.02	0.05
			(0.07)	(0.10)	(0.08)	(0.10)	(0.03)	(0.05)	(0.06)	(0.03)
	Inverted	FS	1.58	1.37	1.20	1.56	0.08	0.18	0.12	0.05
			(0.16)	(0.08)	(0.09)	(0.16)	(0.06)	(0.02)	(0.04)	(0.03)
		FI	1.21	1.03	0.98	1.16	0.13	0.03	-0.14	0.03
			(0.07)	(0.07)	(0.09)	(0.09)	(0.09)	(0.06)	(0.11)	(0.09)
		FO	1.23	0.83	0.77	1.20	0.21	0.22	0.06	0.21
			(0.08)	(0.04)	(0.05)	(0.09)	(0.05)	(0.07)	(0.11)	(0.05)

Note. X_1Y_2 represents the distribution at level 1 in dimension X and level 2 in dimension Y. Dimension X is the top of the face and dimension Y is the bottom of the face, level 1 is same/upright and level 2 is different/odd. FS, Feature Size Task; FI, Feature Identity Task; FO, Feature Orientation Task.

Post-hoc analysis of the Feature Orientation and Feature Identity Tasks using Bonferroni corrected paired *t*-tests revealed that, in the Feature Identity Task, there was an effect of orientation for the top-same-bottom-different distribution but not the others ($t(5) = 8.78$, $p < 0.001$). A positive mean rho value was observed in the inverted condition ($M = 0.07$, $SE = 0.09$) compared to a negative mean rho value in the upright condition ($M = -0.09$, $SE = 0.07$). However, caution must be given in interpreting this contrast as neither correlation is significantly different from zero.

In the Feature Orientation Task, there was an effect of orientation for the eyes-normal-mouth-normal distribution ($t(5) = -7.58$, $p = 0.001$), with a positive mean rho value in the upright condition ($M = 0.46$, $SE = 0.10$) compared to a negative mean rho value in the inverted condition ($M = -0.04$, $SE = 0.09$). In this case, responses to normal eyes and mouths correlate in upright but not inverted faces. Therefore, these data suggest that violations of PI are found for normal faces when upright but not when inverted.

This same pattern of results was not found in the other tasks, which may seem surprising, but the designs of the experiments differed. Thus, a typical, unaltered face was not seen in isolation in either an upright or inverted condition in either the Feature Size or Feature Identity Tasks. Supplementary analysis examined distributions for evidence of an inversion effect if the mean rho value for one orientation does not fall into the confidence interval of the mean rho value in the other orientation distribution. Results highlight that inversion effects for rho values are only found for the Feature Orientation Task (Table 2.4). Evidence of orientation specific, between-feature encoding is reported, but only when feature orientation itself is manipulated.

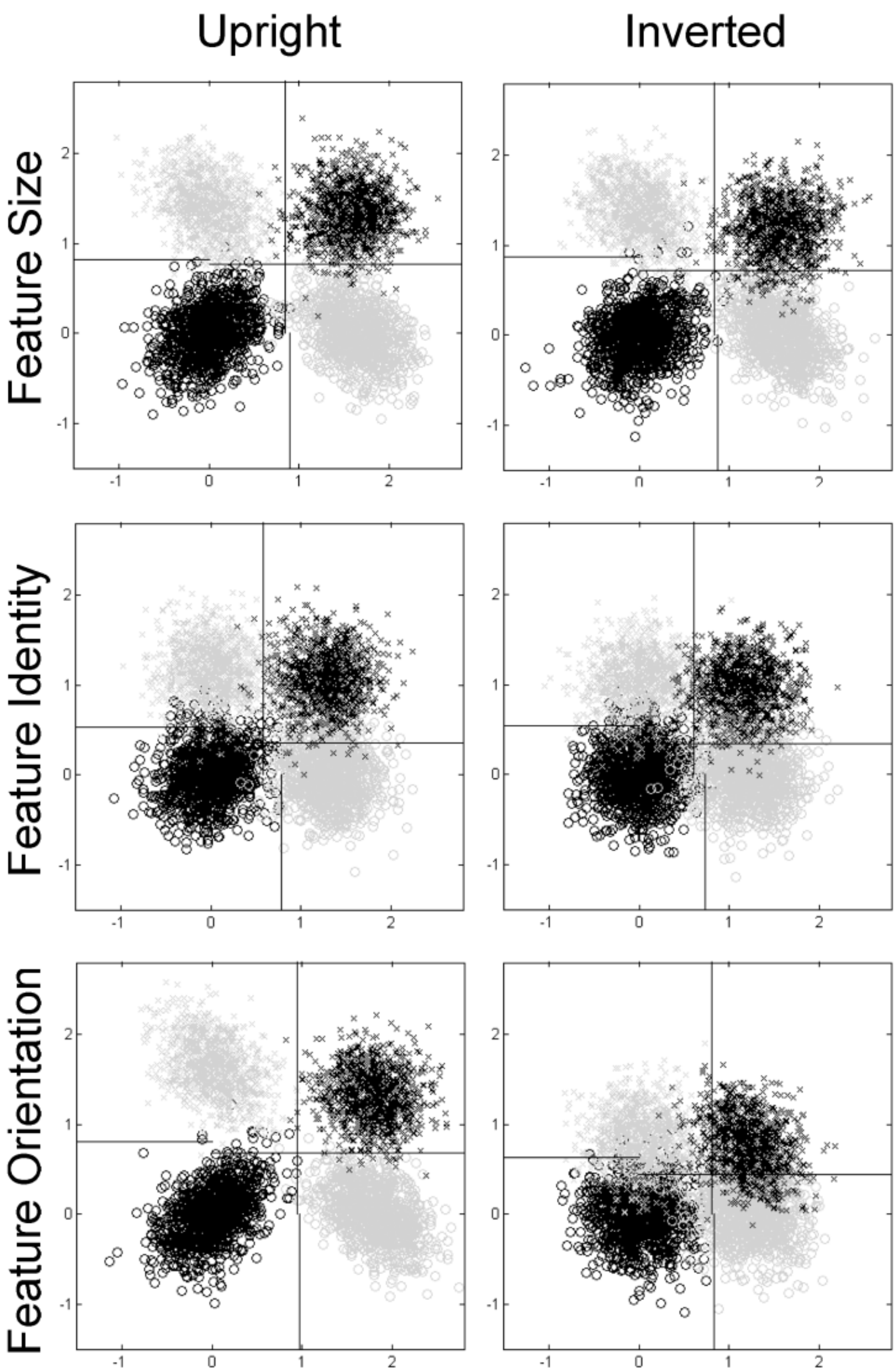


Figure 2.10. Plots of distributions and decision criteria from probit analyses: same (normal) top, same (normal) bottom = black circles; different (odd) top, same (normal) bottom = grey circles; same (normal) top, different (odd) bottom = grey crosses; and different (odd) top, different (odd) bottom = black crosses. Note distributions are plotted relative to an origin at the mean of the bivariate dimensions for the same-same (normal-normal) case. Plots panelled by task (Feature Size, Feature Identity and Feature Orientation) and orientation condition (upright and inverted).

General Discussion

The goal of the present study was to explore face processing in three different tasks. Across all tasks, evidence was sought for a selective marker of whole face processing in upright faces. In the group-based analyses of *d'* and *c*, this marker was defined as an interaction between status of the other feature and orientation. Evidence was found for an interaction between status of the other feature and orientation for sensitivity. However, this evidence was only in the Feature Orientation Task where sensitivity was significantly higher to feature orientation in upright faces when the other feature was 'normal' compared to when the other feature was 'odd'. The orientation-specific effect found in the group-based analysis for the Feature Orientation Task was supported by orientation-specific violations of PS in upright faces, and violations of DS in inverted faces found in the marginal and probit analyses. Also, the significant rho values and inversion effects revealed when data was averaged across participants supported the orientation specificity.

The Feature Orientation Task is an analogue of the Thatcher illusion. The orientation specific violations of PI are limited to the case where features in their normal orientation are presented. In other words, the orientation specific violations of PI do not occur with Thatcherised faces (i.e. upright faces with two inverted (odd) features). This finding is in line with the lack of evidence for excitatory interactions between eyes and mouth for the Thatcher illusion, as determined by processing capacity (RT-based) measures (Donnelly, Cornes, & Menneer, 2012). The absence of within-stimulus dependence between the eyes and mouth in the Thatcher illusion is at odds with the striking and apparently configural phenomenon experienced when viewing the upright versus inverted stimulus. However, one suggestion is that the perception of upright Thatcherised faces is facilitated by socio-emotional encoding mechanisms, which are not activated for inverted Thatcherised faces (see Donnelly et al., 2011). This additional source of socio-emotional information adds to that available in the visual representation to create the phenomenology associated with the illusion.

The Feature Identity task is an analogue of the aligned condition of the composite face task. Previous studies of the composite face effect that have used a feature-complete design also reported a significant role for decisional influences on performance (Richler et al., 2008). Moreover, it is noted that exploring the aligned condition of the composite face task using the feature-complete design is similar to use of the ‘complete design’ (as opposed to the ‘partial design’, see Richler, Cheung, & Gauthier (2011a) for discussion of the differences between partial and complete designs) to explore the impact of alignment and congruency effects on the composite face effect. The complete design includes conditions absent from partial designs, where the non-probed feature is manipulated so that if responses were required for the non-probed feature, correct responses would be either congruent or incongruent with those made to the probed feature. Both feature-complete and complete designs index the influence of response congruence. Explicit in the use of both designs is the idea that measuring congruence effects is important in studies of configural and holistic face processing. The present findings confirm that studies of the composite face effect need to include measures of the effect of feature congruence (status of the other feature) on performance (see also Richler et al., 2011a).

Across Feature Size, Identity and Orientation Tasks one general comment must be made. While evidence of orientation specific violations is consistent with a qualitative effect of orientation on face perception, its absence is not. Therefore, only one task (the Feature Orientation Task) demonstrates a qualitative effect. However, it is important to note that orientation may still exert a quantitative influence on performance. Increased sensitivity or bias across orientation is not evidence for dependencies (configurality) between features, but it is important to capture and measure. The enhanced sensitivity to upright over inverted faces in the Feature Size and Feature Identity Tasks indicates just such an effect. Determination of feature size and identity is better for upright than inverted faces, even though in each task orientation does not change the fundamental influences on performance. This connects these effects in face perception to effects of canonical orientation in object identification and perception (e.g. Palmer, Rosch, & Chase, 1981; Rock, 1973; Tarr & Pinker, 1989). In the Feature Identity Task, the same pattern in bias was present for both upright and inverted composite faces, but the type of violation in DS differed by orientation. This demonstrates that GRT analysis is required to distinguish the nature of the effect, but also that GRT tasks can still replicate behavioural findings.

By using the feature-complete design to explore sources of configurality across a family of related feature manipulations (size, identity and orientation), it has been possible to directly compare how underlying processes determine differences and similarities across pairs of faces and grotesqueness and typicality within faces. As this is the start of this endeavour, I recognise limitations to the conclusions that can currently be drawn. The immediate focus has been on establishing statistical tests of sufficient sensitivity, allied to running participants in experimental conditions very similar to those used previously (Wenger & Ingvalson, 2002, 2003; Richler et al., 2008; Cornes et al., 2011). The consequence is that there are increased processing demands for the Feature Size Task, as it included a distractor task, not present in either the Feature Identity or Feature Orientation Tasks. These task differences may have impacted on the differential findings. Also, the nature of asking participants to make decisions about two features on each trial means that results must be interpreted in a different way to compare with effects in tasks that require only one response. Despite the large number of conditions, it would also be valuable to explore non-face control conditions for the Feature Orientation Task to explore whether the violations of DS seen with inversion are face specific. Furthermore, examining misalignment in the Feature Identity Task could be useful to see if consistent violations of PS emerge.

The current data support the view that the stimulus and task manipulations lead to differences in sources of configurality across tasks. These differences are not readily captured by notions of holistic and second-order relational processing that have been derived for the purpose of explaining face specific processing. Instead, these stimuli and task effects are tested against mathematically defined statistical violations, articulated within the constructs of GRT, which are indicative of dependencies between features. In seeking to establish appropriate tests of these violations, the present study has demonstrated that probit analysis is a useful addition to the set of analytic tools used to draw inferences about configurality within GRT. Multiple converging techniques for relating data to theory can help identify the most likely underlying nature of the effect. Knowledge of the reliability of the methods given the parameters of this study allows informed interpretation of differential inferences. Furthermore, exploring the correlations between the feature dimensions of bivariate distributions allowed inferences about the stimulus conditions that support the encoding of between-feature, within-face relationships. In doing so, both the presence and magnitude of such relationships can be compared. In summary, using probit analysis allowed violations of PI that remained undetected by marginal analysis to be reported.

This study has highlighted some important points that apply to the traditional behavioural approach to investigating configurality. Some participants appear consistently more sensitive to the tasks than others. The Feature Identity Task had lower levels of sensitivity overall and still a high number of violations emerged. This shows low levels of sensitivity do not necessarily lead to fewer violations; hence, reduced evidence for configurality. Also, there are differences between participants in the pattern of violations shown within a task. These violations are based on performance over hundreds of trials, so can be considered very robust. This highlights how averaging or aggregating across participants, even within a robust test, can lose richness of the data to demonstrate differences in configurality between participants.

Conclusion

The data presented are consistent with the determination of feature size, identity and orientation in upright faces being subject to several influences. Results demonstrate that there are differences in sources of perceptual and decisional configurality between these three tasks. By using a GRT methodology, these differences have been revealed. If other tasks were explored in this way then the broad similarities and differences between face processing tasks could be understood. Also, the findings have demonstrated the probit model approach is valuable in revealing violations of PI where marginal analysis fails.

Chapter 3

The Impacts of Thatcherisation and Inversion on the ERP Components of Face Processing

Chapter 2 demonstrated that only the Feature Orientation Task (Thatcher illusion task) generated reliable evidence for all three violations of GRT constructs: perceptual independence (PI), perceptual separability (PS) and decisional separability (DS). The co-occurrence of these violations is consistent with three different influences when the perception of Thatcherised faces is compared between upright and inverted orientations. In Chapter 3, evidence of neural correlates associated with these influences of configural processing in relation to the Thatcher illusion is sought. To understand the spatial and temporal bases of these influences and their neural generators requires a method that can examine the time course of brain responses to face processing in the millisecond (ms) range. The event-related potential (ERP) paradigm allows examination of the neuroelectrophysiology of face processing, including configural face processing.[4]

Many studies have explored the face sensitivity of ERP components. The visual P1 is the first positive component generated in response to visual stimuli, occurring circa 100 ms post stimulus, with the largest amplitude in the occipital region. The P1 is sensitive to low-level visual features such as luminance, contrast and spatial frequencies (Regan, 1989). However, the P1 component has been reported to be larger in response to faces than objects (e.g. Eimer, 2000a). While some have interpreted this enhanced P1 as a marker of face categorisation, recent studies have shown this conclusion is likely to be incorrect. Rousselet, Husk, Bennett and Sekuler (2008) found face-specific effects disappeared when the faces and objects were matched for low-level properties. Therefore, the evidence for a face-specific P1 response is not robust.

---

[4] A basic introduction to the ERP technique and the notation used in this chapter can be found in Appendix C.

The N170 (the face specific N1) is the first negative deflection peaking at around 160-170 ms post stimulus. It is often largest in the occipito-temporal region and is particularly large in response to faces relative to objects (e.g. Rossion & Jacques, 2012). It is found not only in response to faces, but also for Mooney faces (George, Jemel, Fiori, Chaby, & Renault, 2005) and faces made from a collection of objects ('Arcimboldo' faces, Rossion & Jacques, 2008). Although reliably larger for faces compared to objects, there are also differences in the amplitude of N170 between objects (Rossion, Gauthier, Tarr, Despland, Bruyer, Linotte, & Crommelinck, 2000; Itier & Taylor, 2004; Rousselet et al., 2008; Rossion & Jacques, 2008).

The effect of inversion on face recognition has also been well documented in a range of behavioural (e.g. Yin, 1969), neuroimaging (e.g. Haxby, Ungerleider, Clark, Schouten, Hoffman, & Martin., 1999; Haxby, de Haan, & Johnson 2000) and ERP (e.g. Eimer, 2000b; Jacques and Rossion, 2007) studies. Inversion has been shown to increase and delay the N170 for faces and objects (Rossion, Delvenne, Debatisse, Goffaux, Bruyer, Crommelinck, & Guérit, 1999; Bentin, Allison, Puce, Perez, & McCarthy 1996). However, this effect of inversion on amplitude of the N170 is larger for faces than objects (Rossion et al., 2000). Finally, consistent with a face and orientation specific component, the amplitude of the N170 is larger in the right hemisphere compared to the left hemisphere (Rossion & Jacques, 2008). Sensitivity of the right hemisphere to upright faces supports the view that the N170 may be an indicator of perceptually-based configural processing. In support of this notion, it is interesting to note that in at least some cases of apperceptive prosopagnosia the specificity of the N170 for faces is lost (Bentin, Deouell, & Soroker, 1999; Eimer & McCarthy, 1999).

The P2 component is the second positive peak elicited at around 200 ms post stimulus. It is largest in the occiptio-parietal region and centro-frontal regions (see Fonaryova Key, Dova, & Maguire, 2005). The P2 has been shown to be sensitive to manipulations of face structure; amplitude was larger to pairs of faces with spacing rather than featural modifications (Mercure, Dick, & Johnson, 2008), for contrast reversed relative to normal faces (Itier & Taylor, 2002) and reduced for atypical (stretched faces) compared to typical faces (Halit, de Haan, & Johnson, 2000).

The P3b, a subcomponent of the P3 (the third positive peak), is a midline positive deflection that is largest in centro-parietal areas, peaking between 300 and

500 ms post stimulus. Generally, the P3b reflects cognitive processing (Polich, 2012) and response preparation through a 'monitoring' process (Verleger, Jaśkowski, & Wascher, 2005). However, the P3b component is larger for faces than objects (Allison, Puce, Spencer, & McCarthy, 1999), especially for threatening faces (Schupp, Öhman, Junghöfer, Weike, Stockburger, & Hamm, 2004). Whether threat influences the P3b directly or by virtue of an effect on cognitive processing or response preparation is unclear.

The Thatcher Illusion and ERP Components

The brief review of face-related components would predict effects of Thatcherisation of faces (relative to typical faces) might influence the N170, P2 and P3b components, but it is unlikely to influence the P1. Consistent with this, Thatcherising faces (compared to typical faces) reduces the amplitude (Gu, Li, Yang, & Zhu, 2007; Boutsen, Humphreys, Praamstra, & Warbrick, 2006; though see reports of increased amplitude Milivojevic, Clapp, Johnson, & Corballis, 2003 and Carbon, Schweinberger, Kaufmann, & Leder, 2005) and delays the latency of N170 and the P2 (Boutsen et al., 2006). Thatcherisation reduces the amplitude of the P2 (P250) for upright faces but this difference in magnitude gradually reduced with orientation, with no differences observed for P2 amplitude for fully inverted faces (Milivojevic et al., 2003). Milivojevic et al. (2003) explored the parietal component (450-600ms) at P_1 and P_2 electrodes. They found that Thatcherisation leads to increased amplitude and this effect was reduced abruptly with inversion of faces. They said their findings were linked to the P3 component and stimulus recognition difficulties for participants to identify the gender of face stimuli in their study (Milivojevic et al., 2003). No study (that I am aware of) has specifically explored the effect of Thatcherising faces on the P3b in relation to decisional response related processes. Most studies confirm that there are no effects of Thatcherising faces on P1 (e.g. Carbon et al., 2005; Boutsen et al., 2006; Gu et al., 2007, although see Milivojevic et al., 2003). Therefore, the working hypothesis is that perceptual violations may influence the N170 and P2 while decisional violations might influence the P3b.

Experiment 3.1

In Experiment 3.1 evidence is sought to see whether perceptual factors influence the N170 and P2 while decisional factors influence the P3b. Previous ERP studies of the Thatcher illusion (and in fact many behavioural and neuroimaging studies bar Cornes, Donnelly, Godwin, & Wenger, 2011; Donnelly et al., 2011; Donnelly, Cornes, & Menneer, 2012) assume that the effect of inverting eyes and mouths is 'as one' on undifferentiated representations of faces. Such a view runs counter to the data of Experiment 2.1, as well as Cornes et al. (2011) and Donnelly et al. (2011), Donnelly et al. (2012). The fact that the different types of stimulus are created by orthogonal manipulations of upright and inverted, eyes and mouths is of value to the present study. Beyond face orientation, the four types of face provide granularity for differentiating one source of influence on ERP from another. The present study explores inversion and levels of Thatcherisation on the P1, N170, P2 and P3b by comparing upright and inverted faces with four levels of Thatcherisation: Thatcherising faces only at the eyes or mouths, relative to both features Thatcherised or neither.

Previous studies expected a purely perceptual locus of the effects; electrocephalogram (EEG) recordings were typically focused to particular posterior sites. Given evidence for a decisional nature to some influences on configurality, it may be that this focus has been incorrect. To avoid missing potentially important effects, a data driven method of analysis has been adopted to locate clusters of electrodes to analyse.

Aims

(1) To examine evidence for effects of inversion and Thatcherisation on P1, N170, P2 and P3b.

(2) To examine whether the level of manipulation (one or two Thatcherised features) is related to the amplitude and latency of each component.

(3) To determine whether there is evidence for markers of configurality through mapping of findings on to violations of GRT constructs.

Method

Participants

Twenty-three participants were recruited at the University of Southampton and participated in the study in return for payment. There were eight males and 15 females with an age range of 19-43 (*M* = 25.52, *SD* = 5.25). Participants were right handed with normal or corrected-to-normal vision. Participants completed an exclusion questionnaire (Appendix D) prior to participation and were selected as suitable to take part. Only right-handed non-smokers with no history of epilepsy were selected. Also, participants who were taking or had recently taken psychoactive medication were excluded from the study. Sixteen participants were included in the final sample (seven males and nine females with an age range of 20-43, *M* = 26.13, *SD* = 5.86); two were excluded for poor task accuracy compared with the rest of the sample (Participants 15 (72.55%) and 23 (74.27%)), one for poor impedance values (Participant 3) and four for poor quality data as fewer than 40 trials per condition remained after data had been through the pre-processing stages (Participants 9, 10, 17 and 19).

Design

A two (orientation: upright or inverted) x four (stimuli manipulation: typical face with normal eyes and normal mouth (NN), eyes Thatcherised (TN), mouth Thatcherised (NT) or both eyes and mouth Thatcherised (TT)) repeated measures design was used. Each participant completed 1920 trials, 240 for each of the eight conditions outlined above. The task was split into 16 blocks of 120 trials and trials were randomised.

Stimuli

Forty upright faces from the set of Thatcherised stimuli created for use in Experiment 2.1 were selected for use in this experiment. The stimuli set consisted of 10 identities (NimStim face stimuli, Tottenham et al., 2009) with the four possible stimuli types outlined above. All 40 images were converted to grey-scale and controlled for low-level properties, similar to Rouselett et al. (2008). A two dimensional fast Fourier transform was computed to equate the stimuli in terms of spatial frequency content by taking the average of the amplitude spectra of all 40 stimuli and using this to reconstruct each of the images with the original phase information. The mean RMS contrast value over the images was 0.22 and the standard deviation was 0.01. A Butterworth filter was used to remove the influence of the face outline. The images were grey-scale faces within an oval annulus and were presented on a black background (Figure 3.1). These images were also rotated by 180° (i.e. inverted) to create a full set of 80 images. The images appeared individually in the centre of the screen at a size of 6.00 cm by 8.50 cm and at a visual angle of 3.44° by 4.80° when viewed from a distance of 100cm, maintained by the distance of the experimental chair from the screen.

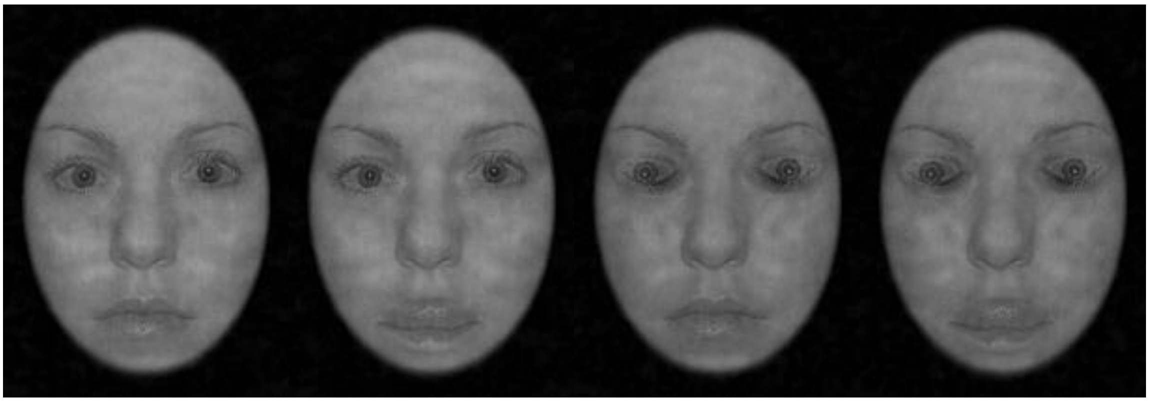


Figure 3.1. Face stimuli used in Experiment 3.1. Conditions NN, TN, NT and TT presented left to right.

Apparatus and Materials

The EEG from the experimental session was recorded using SCAN 4.4™ (copyright © 2006, Compumedics Neuroscan). The experiment was created in Presentation 14.9 (copyright © 2003-2010, Neurobehavioral Systems) and this software allowed markers for display events to be recorded into Aquire for SCAN 4.4™ to time lock the stimulus onset to the recoding. The low-level control to stimuli was altered using MATLAB® (2009a, The MathWorks, USA). The Edinburgh Handedness Inventory (Oldfield, 1971) and the exclusion questionnaire (see Appendix D) were used to identify participants that would not be suitable to take part. All stimuli were presented in a darkened room on a desktop computer with a screen size of 32.80 cm by 24.50 cm at a distance of 100 cm from eye level to the screen. Screen resolution was 1024 by 768 and refresh rate of 85 Hz. Participants responded by clicking buttons on a two button response box. All stimuli were presented against a black background and text instructions were presented in white.

Procedure

Participants were seated in a dark room and maintained their head in an upright position. Participants were shown the face stimuli and asked to make a speeded judgement as to whether the face was upright or inverted. On each trial, a blank black screen was presented for about 200 ms, followed by a small white fixation cross in the middle of the screen for a random duration ranging from 500 to 900 ms. A stimulus was then presented centrally for 100 ms, followed by a blank screen for 1000 ms during which time participants responded using a button press. If no response was made, then this trial was considered an error (incorrect). Trial durations thus ranged from 1800-2200 ms. Participants were given the opportunity to take breaks between blocks.

EEG Recording and Pre-Processing

EEG recording

The EEG data were acquired from 60 channels (see Figure 3.3) using a 10-20 system Easycap (Brain Products GmbH) and a SynAmps^2 amplifier headbox (Neuroscan, Compumedics) with silver/silver chloride (Ag/Ag Cl) electrodes. The vertical EOG was monitored from an electrode above the right eye against an electrode below the right eye. Analogue signal was digitalised at 500 Hz and band-pass filtered between 0.1 and 100 Hz. The ground electrode was located along the midline at FC_z and the active electrode was located at AF_z. Impedances were kept below 10 kΩ and were typically below 5 kΩ. Participants were asked to keep movement to a minimum during the trials. They were also asked to minimise blinking, swallowing and tension.

EEG pre-processing

Pre-processing was performed using the EDIT menu in SCAN 4.4™. The EEG data were re-referenced offline to an average reference. Electrodes were rejected on a participant-by-participant basis. Two participants had electrodes excluded from analysis due to imprudence exceeding 10 kΩ. Participant 5 had electrode FP_1 excluded and Participant 6 had electrodes FP_2 and AF_5 excluded. This gave a minimum number of 58 electrodes. The signal was low-pass filtered at 30 Hz (48 dB/octave). Baseline correction was performed using 300 ms of pre-stimulus activity. Eye movements were corrected using the ocular artefact reduction algorithm within SCAN 4.4™ for the VEOG channel (Semlitsch, Anderer, Schuster, & Presslich, 1986). Artefacts were rejected based on absolute values larger than 100 μV. Only correct trials were averaged using an interval from -300 ms to +800 ms. Across the eight conditions, the minimum number of trials was 1955, the maximum 2097 and the mean 2025 (out of a possible 3840). Across participants, the minimum number of trials contributing viable data to a condition was 40, the maximum 208 and the average number of epochs for each condition was 126.56. A repeated measures ANOVA revealed no significant differences between conditions in terms of the number of trials contributed by each participant ($F(7,105) = 1.18$, $p = 0.322$). GFP was calculated for each participant in each of the eight conditions, and averaged to create a grand average global field power (GFP, Lehmann & Skrandies, 1980) for all participants across the whole task.

Results

Performance/Behavioural Data

Accuracy across all trials was calculated prior to pre-processing. Average accuracy and reaction times (RTs) across the conditions are shown in Table 3.1 for the participants included in the analysis. A repeated measures factorial ANOVA was conducted on the RTs from correct trials in each condition from the 16 participants that were used in the final peak detection analysis (see participants section in method for participant exclusion details). The ANOVA factors were orientation (upright and inverted) and condition (normal eyes and normal mouth (NN), normal eyes and mouth Thatcherised (NT), eyes Thatcherised and normal mouth (TN), both eyes and mouth Thatcherised (TT)). Normality was checked using the Shapiro-Wilk test as samples across the conditions in each of the tests were less than 50. The statistic did not reach significance in any of the eight conditions ($ps > 0.056$). When Mauchly's Test of Sphericity was violated, the Greenhouse-Geisser adjusted degrees of freedom are reported. The main effect of orientation was significant ($F(1,15) = 7.99$, $p = 0.013$, $\eta_p^2 = 0.35$). RTs were faster to upright ($M = 613.53$ ms, $SE = 19.92$) than inverted ($M = 633.13$ ms, $SE = 23.90$) faces. Neither the main effect of condition nor the interaction of orientation and condition reached significance ($F(1.69,25.39) = 2.78$, $p = 0.089$, $\eta_p^2 = 0.156$; $F(1.84,27.53) = 0.72$, $p = 0.483$, $\eta_p^2 = 0.046$). In summary, the behavioural data show that participants were able to accurately perform the task by responding whether the faces were upright or inverted, and upright faces were responded to faster than inverted faces.

Table 3.1

Mean Accuracy and Mean Reaction Time across the Upright and Inverted Face Conditions

	Upright				Inverted			
	NN	NT	TN	TT	NN	NT	TN	TT
Accuracy	96.12	94.77	95.70	94.11	94.51	94.06	94.87	93.46
(%)	(2.14)	(2.32)	(2.51)	(3.56)	(2.41)	(3.46)	(2.72)	(2.87)
Reaction	601.44	610.49	604.55	611.51	625.21	629.68	624.89	627.19
Time (ms)	(80.23)	(83.48)	(83.29)	(88.02)	(94.27)	(98.68)	(93.65)	(94.46)

Note. *SD* presented in brackets. N = 16.

Electrophysiology

Global field power (GFP) was used as a global description of the data. GFP is the standard deviation across electrodes computed at each time point. The mean GFP across all participants and in all conditions across the epoch and a butterfly plot showing the mean amplitude for each electrode across the epoch are displayed in Figure 3.2. The butterfly plot shows peaks of activity across electrodes corresponding to the peaks in the GFP graph. These peaks of activity represent the components; of interest in this study; the P1, N170, P2 and P3b components.

GFP peaks were not considered appropriate to identify the time windows for analysis of the components; this is as a result of components causing diffuse activation across the scalp that varies in latency. Instead, regions of interest were selected based visual inspection of maximum amplitudes. The corresponding electrodes from the opposite hemisphere were also included to allow comparison of effects by hemisphere as previous studies have shown activity to faces compared to objects is greater in the right fusiform gyrus (Haxby et al., 1999; Kanwisher, McDermott, & Chun, 1997; McCarthy, Puce, Gore, & Allison, 1997, Figure 3.3). For the early components (P1, N170 and P2), the posterior parietal, temporal and occipital regions were examined to find the peak for the components from which to create the windows for peak detection (e.g. Milivojevic et al., 2003; Carbon et al., 2005; Boutsen et al., 2006). For the P1 component the electrodes chosen were I_z, O_z, O_1, PO_7, O_2 and PO_8 (Figure 3.4). The electrode with maximum amplitude at P1 was I_z which peaked at 138 ms. For the N170 component the electrodes chosen were P_7, PO_7, TP_7, P_8, PO_8 and TP_8 (Figure 3.5). The electrode with maximum amplitude at N170 was TP_8 which peaked at 186 ms. For the P2 component the electrodes chosen were I_z, O_z, O_1, PO_7, O_2 and PO_8 (Figure 3.6). The electrode with maximum amplitude at P2 was O_z which peaked at 244 ms. For the P3b component, the central parietal region was selected (Polich, 2012) and the electrodes chosen were CP_z, P_z, CP_1, P_1, CP_2, and P_2 (Figure 3.7). The electrode with maximum amplitude at P3b was CP_z which peaked at 418 ms. A topographical map (Figure 3.8) displays the areas of maximum activity across the epoch at the peaks of maximum amplitudes outlined.

The latency of the maximum peak was used to create the window for peak detection for each component (e.g. Picton et al., 2000). The component windows for the early peaks were 10 samples from this chosen latency (20 ms either side). This size window was chosen as it estimates roughly the top third of the peak. This meant that the P1 window was 118-158 ms, the N170 window was 166-206 ms and the P2 window was 226-266 ms. The component window for the P3b was much larger due to the broader waveform of the ERP peaks in this area of interest. Therefore, 50 samples (100 ms) from the chosen latency appeared to capture the top third of the peak for CP_z (the electrode of maximum amplitude) by visual inspection. The P3b window was 318-518 ms.

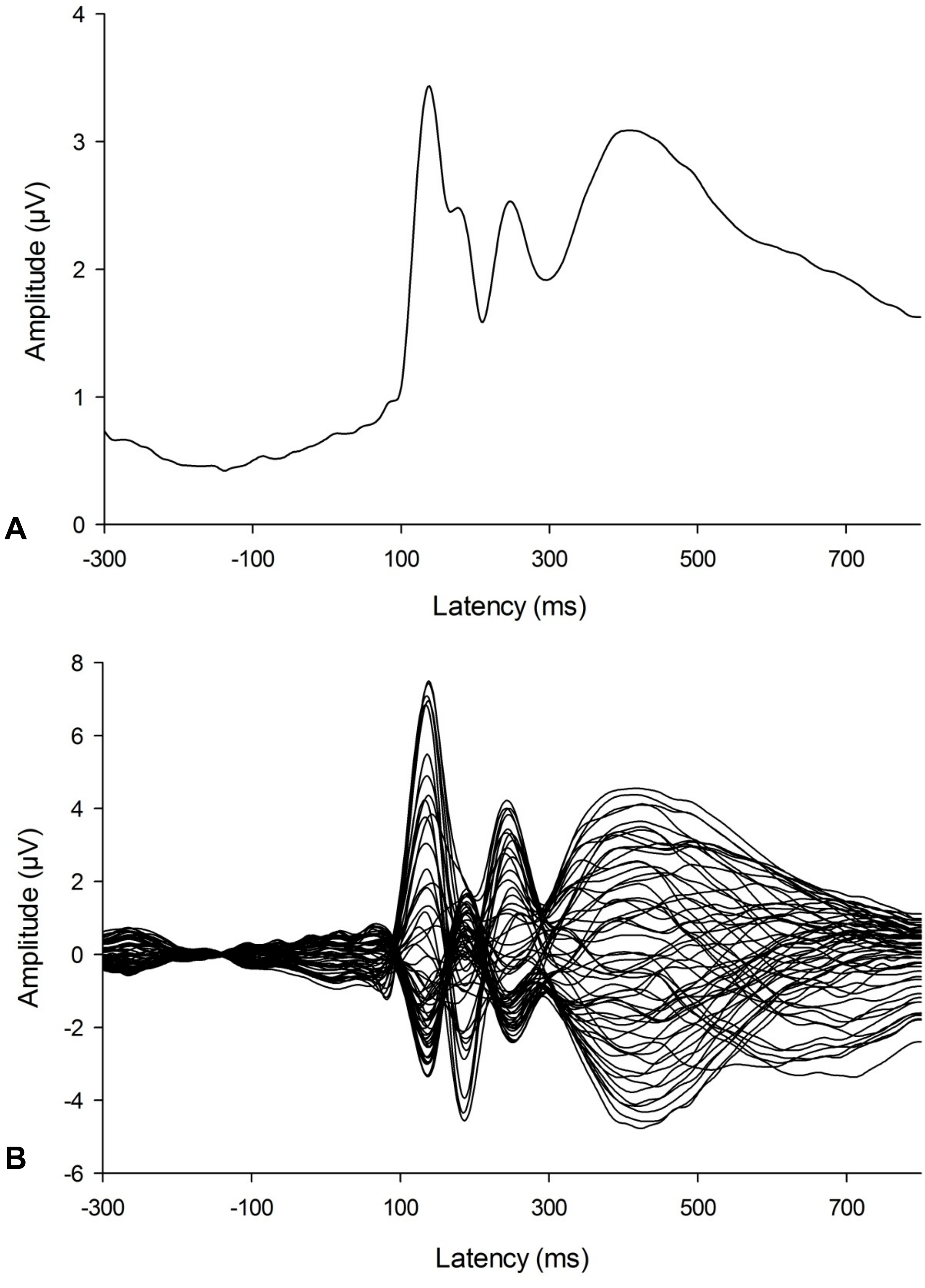


Figure 3.2. (A) Mean Global Field Power (µV) across participants and across all face conditions in the epoch (N = 16). (B) Butterfly plot demonstrating mean amplitudes (µV) for each electrode epoch. Amplitudes averaged across all participants and all face conditions.

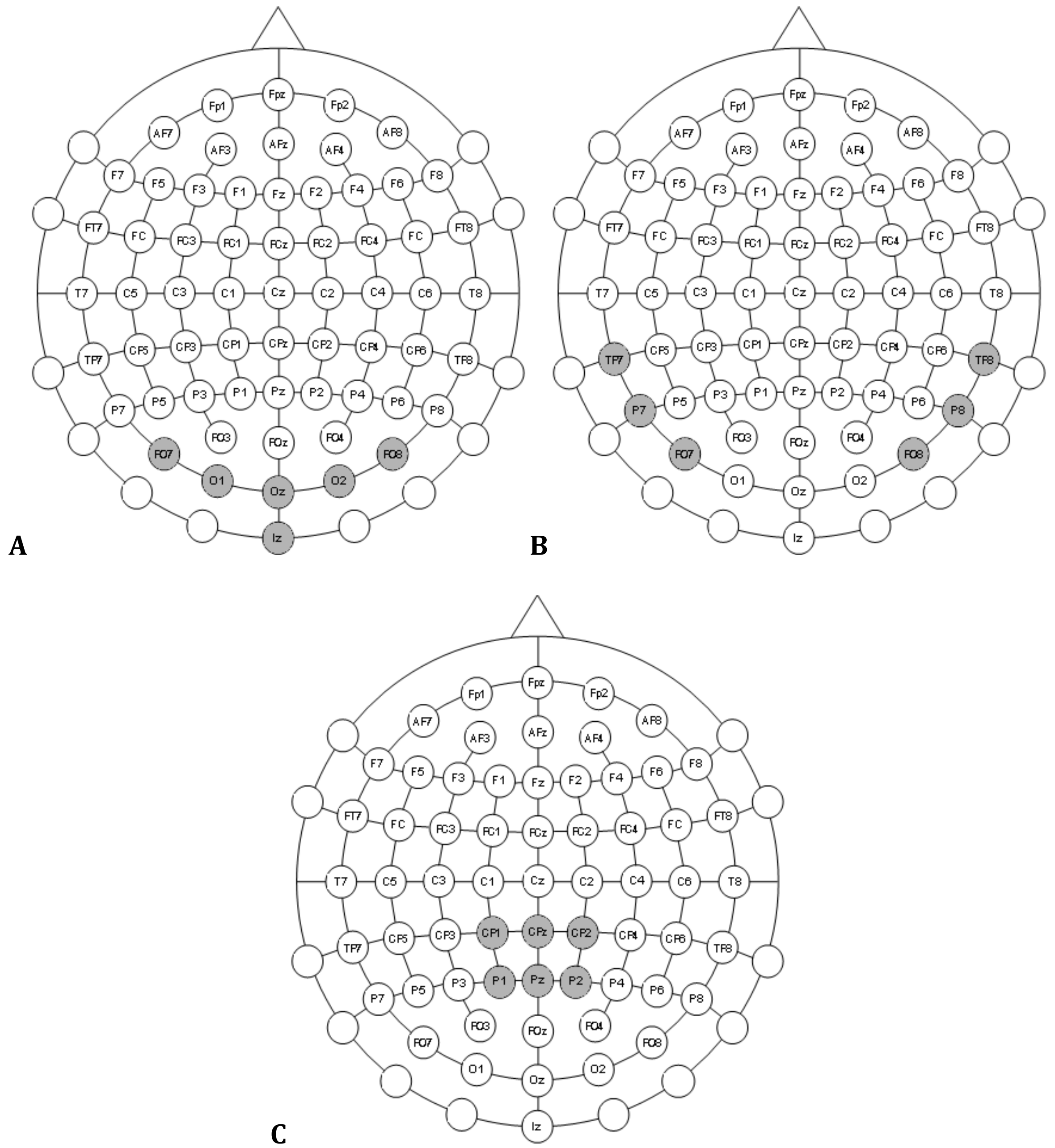


Figure 3.3. Maps showing locations of all electrodes used (those named only). The active electrode was placed at AF_z and ground at FC_z. (A) Electrodes of interest for peak detection of P1 and P2 components highlighted in grey. (B) Electrodes of interest for peak detection of N170 component highlighted in grey. (C) Electrodes of interest for peak detection of P3b component highlighted in grey.

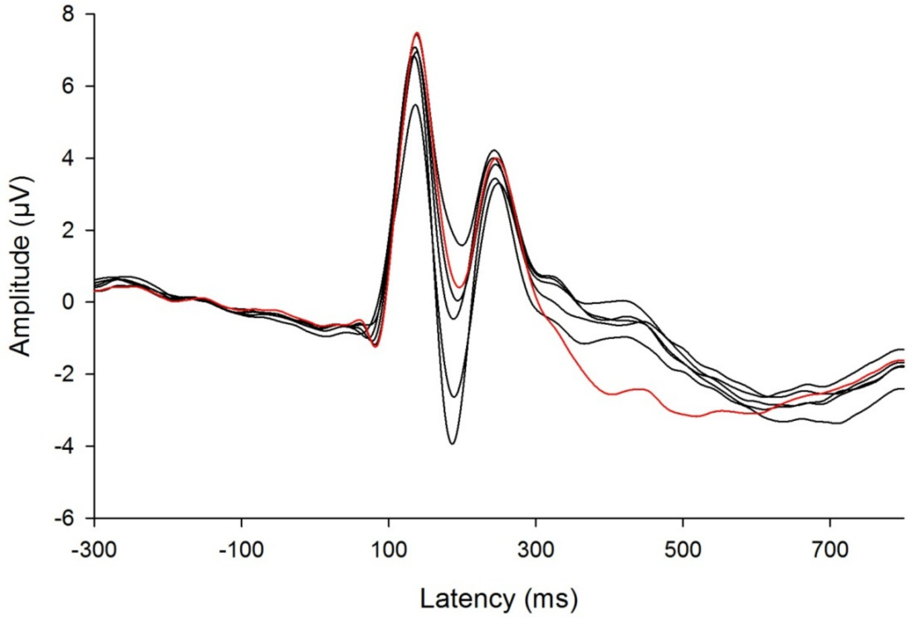


Figure 3.4. Butterfly plot (black lines) demonstrating mean amplitudes (µV) from each electrode of interest (I_z, O_z, O_1, PO_7, O_2 and PO_8) for the P1 component. The red line represents the electrode with the greatest peak amplitude at P1 (I_z), used to define the latency at the centre of the P1 peak window (138 ms).

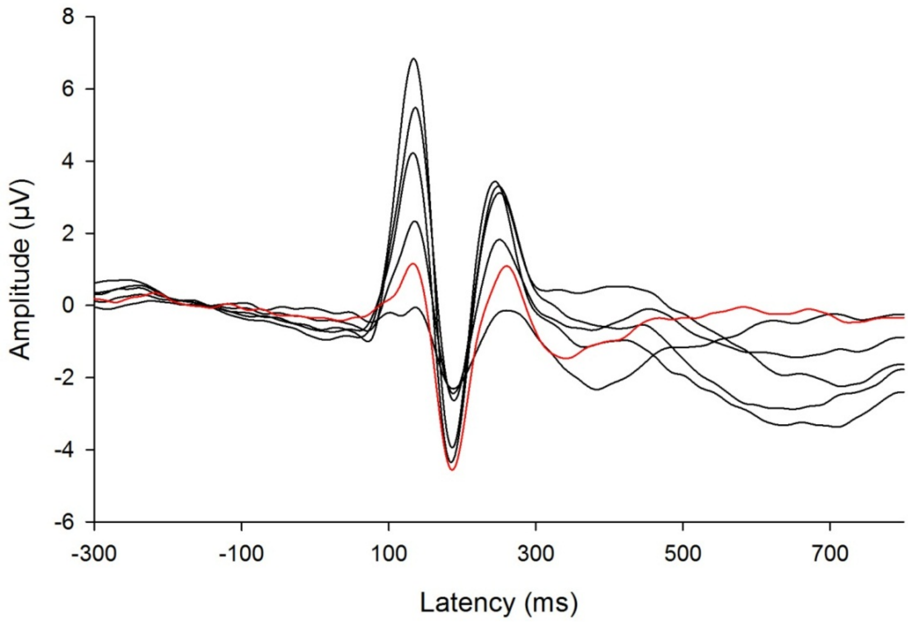


Figure 3.5. Butterfly plot (black lines) demonstrating mean amplitudes (µV) from each electrode of interest (P_7, PO_7, TP_7, P_8, PO_8 and TP_8) for the N170 component. The red line represents the electrode with the greatest peak amplitude at N170 (TP_8), used to define the latency at the centre of the N170 peak window (186 ms).

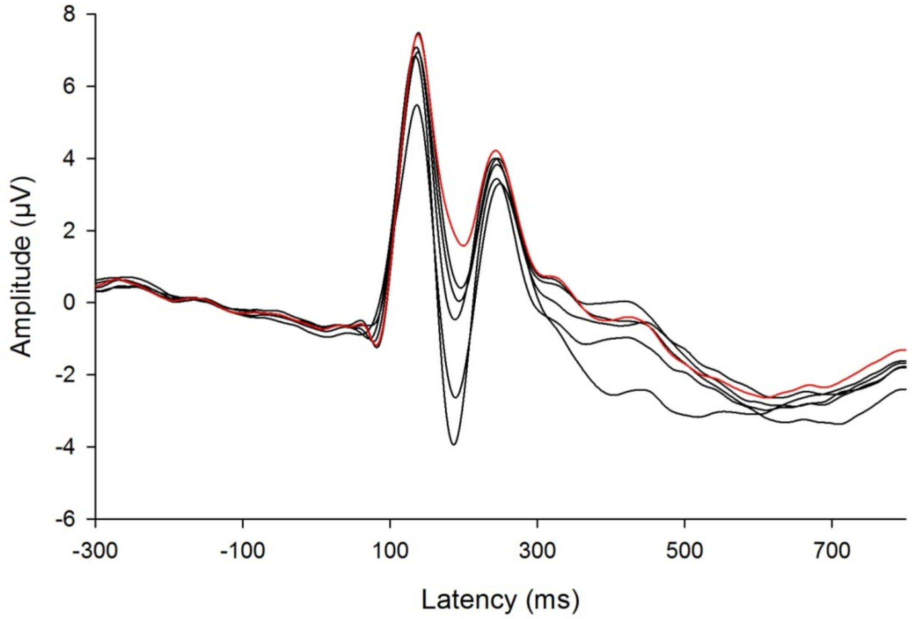


Figure 3.6. Butterfly plot (black lines) demonstrating mean amplitudes (µV) from each electrode of interest (I_z, O_z, O_1, PO_7, O_2 and PO_8) for the P2 component. The red line represents the electrode with the greatest peak amplitude at P2 (O_z), used to define the latency at the centre of the P2 peak window (244 ms).

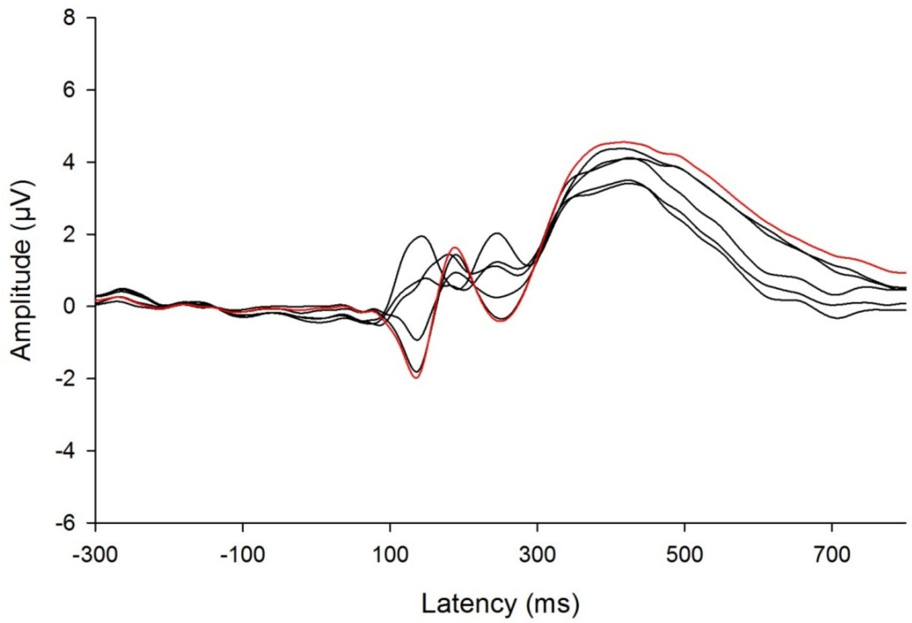


Figure 3.7. Butterfly plot (black lines) demonstrating mean amplitudes (µV) from each electrode of interest (CP_z, P_z, CP_1, P_1, CP_2, and P_2) for the P3b component. The red line represents the electrode with the greatest peak amplitude at P3b (CP_z), used to define the latency at the centre of the P3b peak window (418 ms).

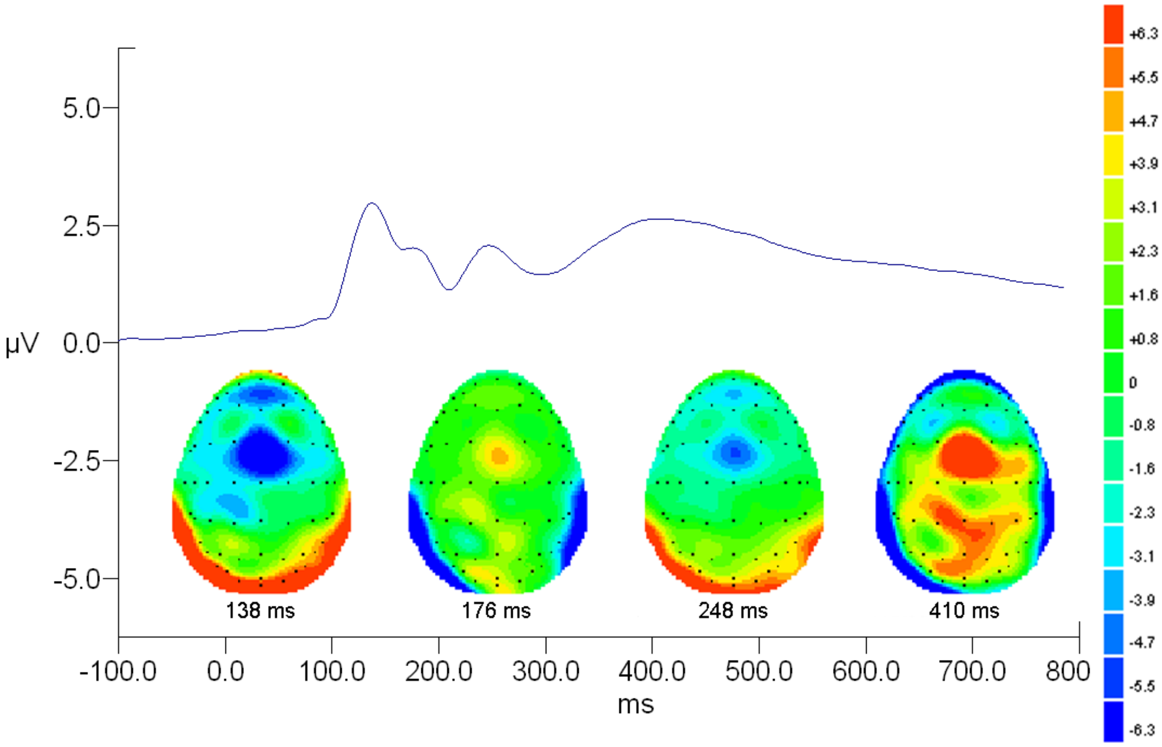


Figure 3.8. A topographical map displaying the areas of maximal amplitude at each of the time points of activity identified.

Event-related Potential Component Analyses

Analysis of peak amplitude and peak latency at each time window of interest for each component used repeated measures ANOVAs. As more than one electrode represented each cluster of interest (e.g., TP_8, P_8 and PO_8 representing the right hemisphere at N170), the peak values (amplitude or latency) were averaged (mean) across electrodes within the electrodes forming each cluster (see Milivojevic et al., 2003).

For the P1, P2 and P3b components, a three (hemisphere: left, midline and right) x two (orientation: upright or inverted) x four (stimulus manipulation: normal eyes and mouth (NN), mouth Thatcherised (NT), eyes Thatcherised (TN) or both eyes and mouth Thatcherised (TT)) repeated measures ANOVA model was used. For the N170 component, the same ANOVA design was used except the hemisphere factor only had two levels: left and right. When Mauchly's Test of Sphericity was violated, the Greenhouse-Geisser adjusted degrees of freedom have been reported. Pairwise comparisons with Bonferroni correction were used to identify differences between levels for the main effects, and simple main effects were used to interpret the interactions. Normality was checked using the Shapiro-Wilk test as samples across the conditions in each of the tests were less than 50. The statistic only reached significance (*p*s < 0.05) in 14 of 176 possible cases. A summary of the ANOVA results are provided in Table 3.2.

Table 3.2

Summary of Findings from the Repeated Measures ANOVAs at each Component of Interest

	P100		N170		P200		P300	
	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Latency (ms)	Amplitude (μV)	Latency (ms)
Hemisphere (H)	ns	ns	ns	ns	ns	*	*	ns
Orientation (O)	ns	ns	*	*	*	ns	*	ns
Condition (C)	ns	ns	ns	ns	ns	ns	ns	ns
H x O	ns	ns	ns	ns	ns	ns	ns	ns
H x C	ns	ns	*	ns	ns	ns	ns	ns
O x C	ns	ns	ns	ns	ns	ns	ns	ns
H x O x C	ns	ns	ns	ns	ns	ns	ns	ns

Note. * indicates a significant effect for the factor at p < 0.05, ns means non-significant.

P1 Component

Peak Amplitude and Peak latency. No main effects or interactions were significant for amplitude or latency (Fs < 2.19, ps > 0.130; Fs < 2.78, ps > 0.078 respectively).

N170 Component

Peak Amplitude. The main effect of orientation was significant ($F(1,15)$ = 8.29, p = 0.012, η_p^2 = 0.35) but the main effects of hemisphere and condition failed to reach significance (Fs < 3.90, ps > 0.067). There was a larger amplitude deflection for inverted (M = -4.41, SE = 0.82) than upright faces (M = -3.84, SE = 0.75).

The interaction of hemisphere and condition was significant ($F(3,45)$ = 4.85, p = 0.005, η_p^2 = 0.24, see Figure 3.9). Simple main effects revealed the condition factor was significant for the right hemisphere, but not the left hemisphere ($F(3,45)$ = 2.92, p = 0.044, η_p^2 = 0.16; $F(3,45)$ = 1.47, p = 0.236 respectively). For the right hemisphere, no pairwise comparisons reached significance; however, there was a significant linear trend across conditions ($F(1,15)$ = 9.21, p = 0.008, η_p^2 = 0.38). The NN condition produced the largest peak amplitude (M = -5.27, SE = 1.15), followed by TN (M = -4.96, *SE* = 1.16), NT (*M* = -4.96, *SE* = 1.18) and TT conditions (*M* = -4.83, *SE* = 1.10). No other interactions reached significance (*F*s < 2.09, *p*s > 0.142).

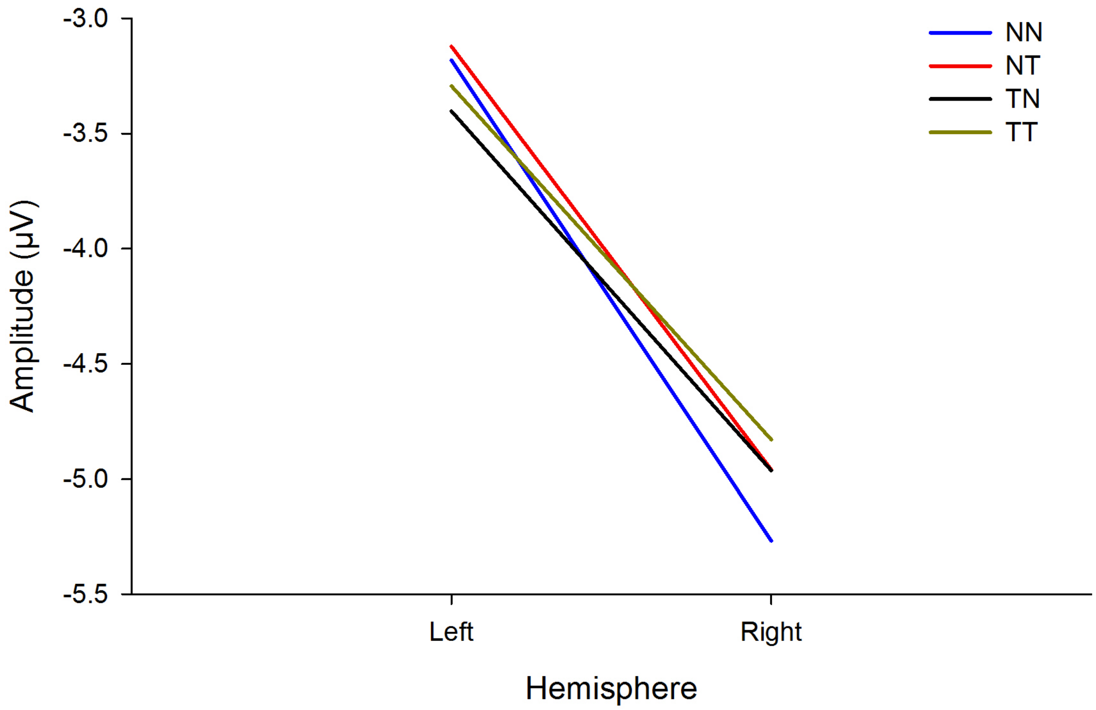


Figure 3.9. Graphs showing the interaction of hemisphere and condition for amplitude of the N170 component.

Peak latency. The main effect of orientation was significant ($F(1,15) = 27.55$, $p < 0.001$, $\eta_p^2 = 0.65$). The peak within the N170 component time window was later for inverted (*M* = 190.37, *SE* = 1.99) than upright (*M* = 185.77, *SE* = 2.06) faces. No other main effects or interactions reached significance (*F*s < 1.71, *p*s > 0.207).

P2 Component

Peak Amplitude. The main effect of orientation was significant ($F(1,15) = 10.73$, $p = 0.005$, $\eta_p^2 = 0.42$). The peak amplitude was larger for upright (*M* = 4.69, *SE* = 0.58) than inverted (*M* = 3.89, *SE* = 0.68) faces. No other main effects or interactions reached significance (*F*s < 2.61, *p*s > 0.122).

Peak latency. The main effect of hemisphere was significant ($F(2,30) = 4.20$, $p = 0.025$, $\eta_p^2 = 0.42$). Pairwise comparisons revealed no significant differences between the latencies of the different electrode clusters, but there was a significant linear trend ($F(1,15) = 6.34$, $p = 0.024$, $\eta_p^2 = 0.30$). Peaks were faster for the right hemisphere cluster ($M = 244.83$, $SE = 2.11$), followed by the midline ($M = 246.32$, $SE = 2.31$) and left hemisphere clusters ($M = 248.32$, $SE = 2.47$). No other main effects or interactions reached significance (Fs < 2.51, ps > 0.120).

P3b Component

Peak Amplitude. The main effects of hemisphere and orientation were significant ($F(1.20,18.05) = 7.42$, $p = 0.011$, $\eta_p^2 = 0.33$; $F(1,15) = 12.09$, $p = 0.003$, $\eta_p^2 = 0.45$ respectively). With respect to hemisphere, pairwise comparisons revealed significant differences between the latencies of the midline ($M = 5.24$, $SE = 0.39$) and left hemisphere ($M = 4.49$, $SE = 0.32$) electrode clusters ($p = 0.001$), midline and the right hemisphere ($M = 4.72$, $SE = 0.46$) electrode clusters ($p = 0.009$) but not between the right and left hemisphere clusters ($p = 1.000$). With respect to orientation, peak amplitude was greater for inverted ($M = 5.13$, $SE = 0.43$) than upright ($M = 4.50$, $SE = 0.34$) faces. No other main effects or interactions reached significance (Fs < 2.04, ps > 0.080).

Peak latency. No main effects or interactions reached significance (Fs < 1.29, ps > 0.271).

The ERP waveforms are plotted for the PO_8 electrode as an example for effects in the right hemisphere for the P1, N170 and P2 and the CP_z electrode for midline effects at P3b. The waveforms are plotted across all upright compared to all inverted conditions (Figure 3.10 and Figure 3.11), across the four upright conditions (Figure 3.12 and Figure 3.13) and across inverted conditions (Figure 3.14 and Figure 3.15).

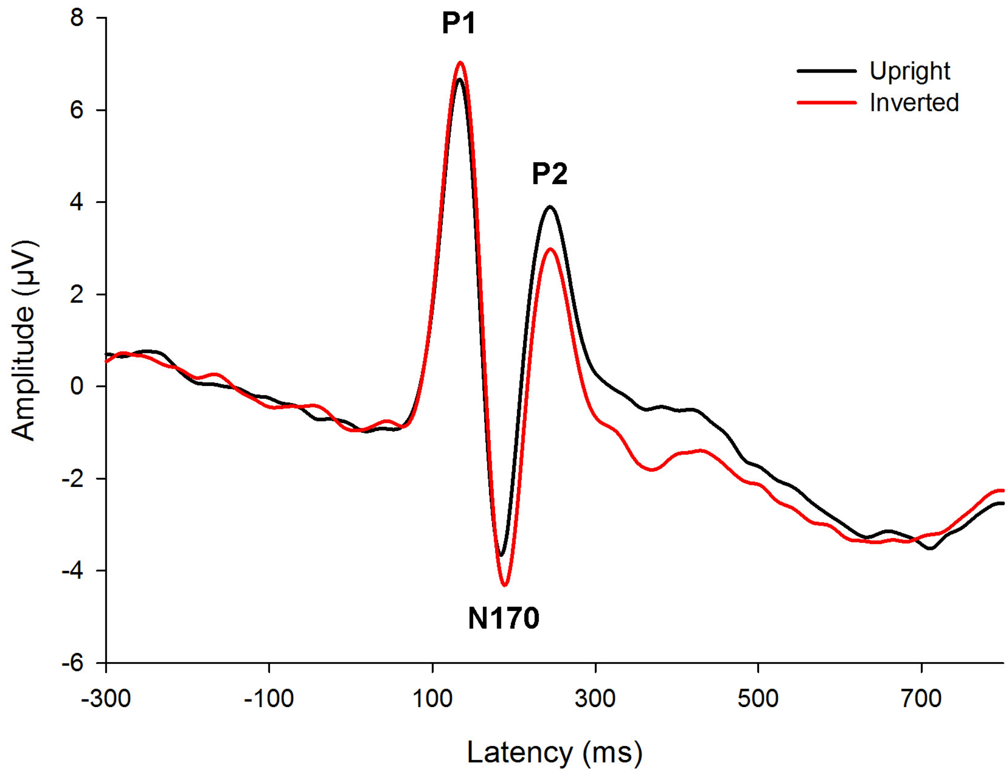


Figure 3.10. Amplitude (µV) across upright and inverted face conditions at the PO_8 electrode.

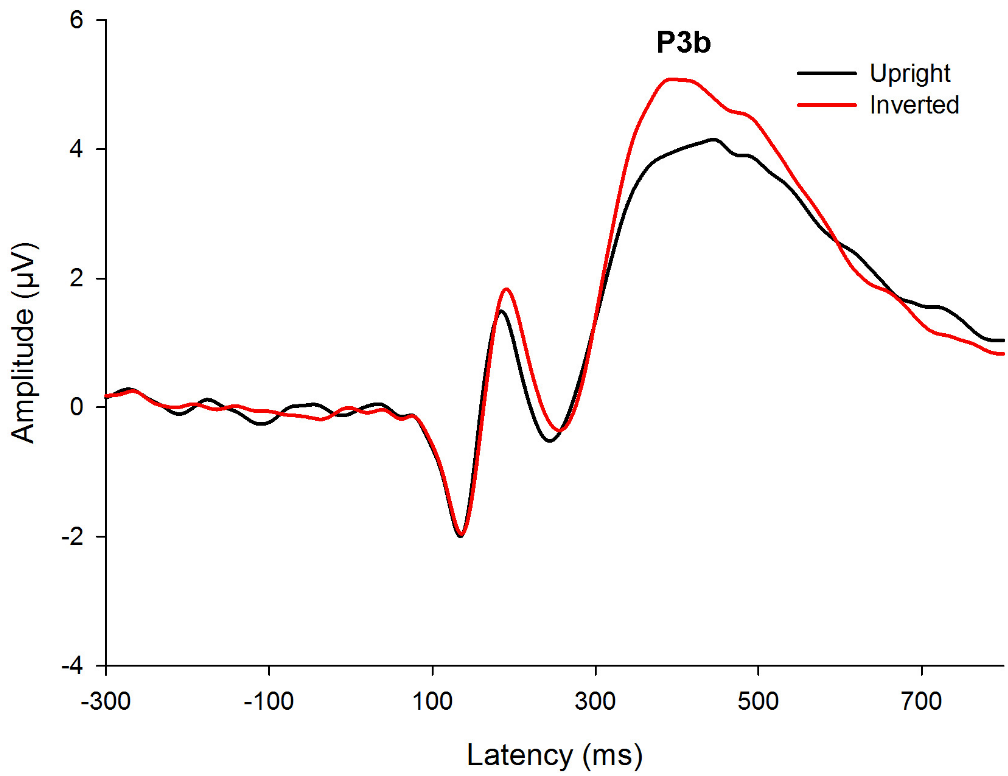


Figure 3.11. Amplitude (µV) across upright and inverted face conditions at the CP_z electrode.

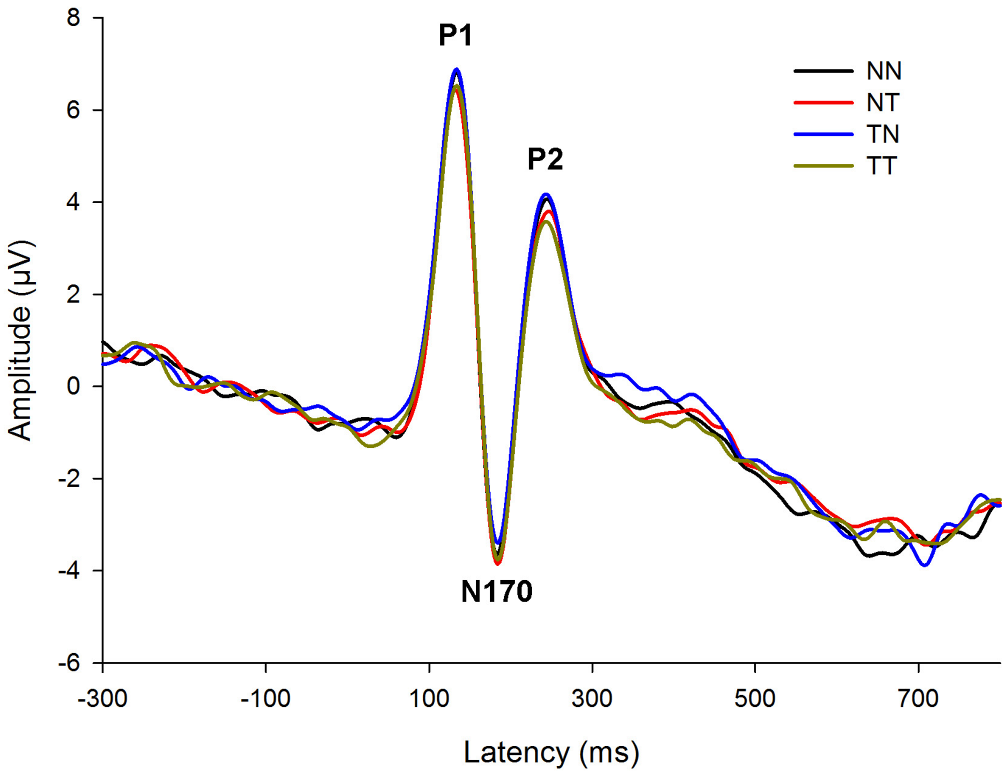


Figure 3.12. Amplitude (µV) across each upright face condition at the PO_8 electrode.

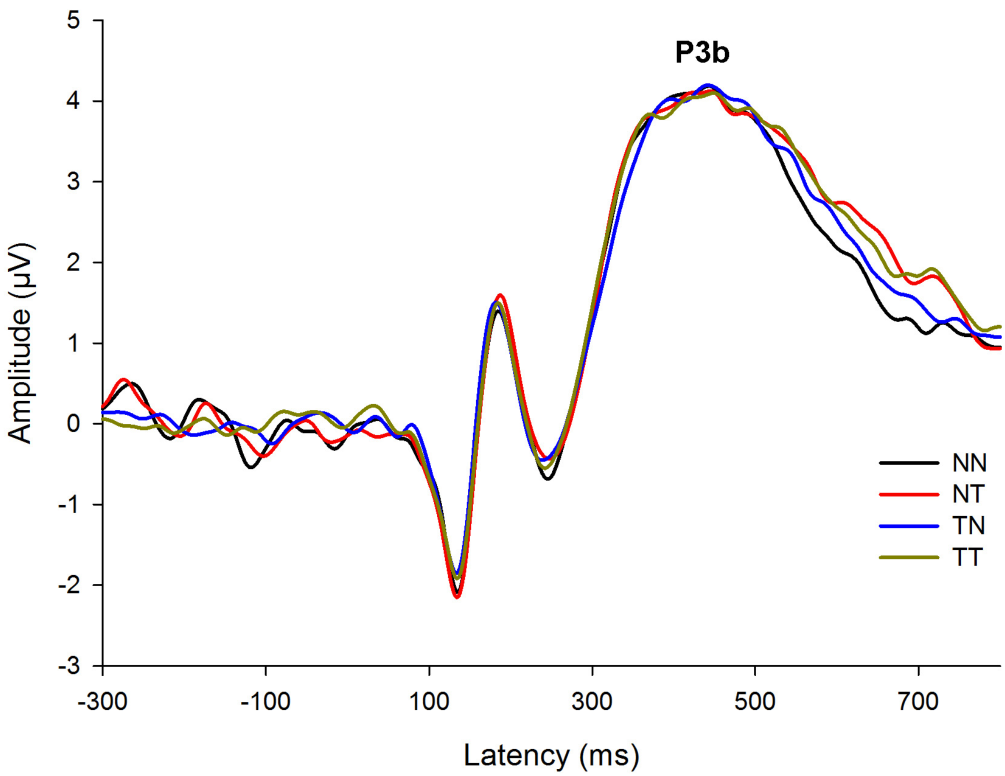


Figure 3.13. Amplitude (µV) across each upright face condition at the CP_z electrode.

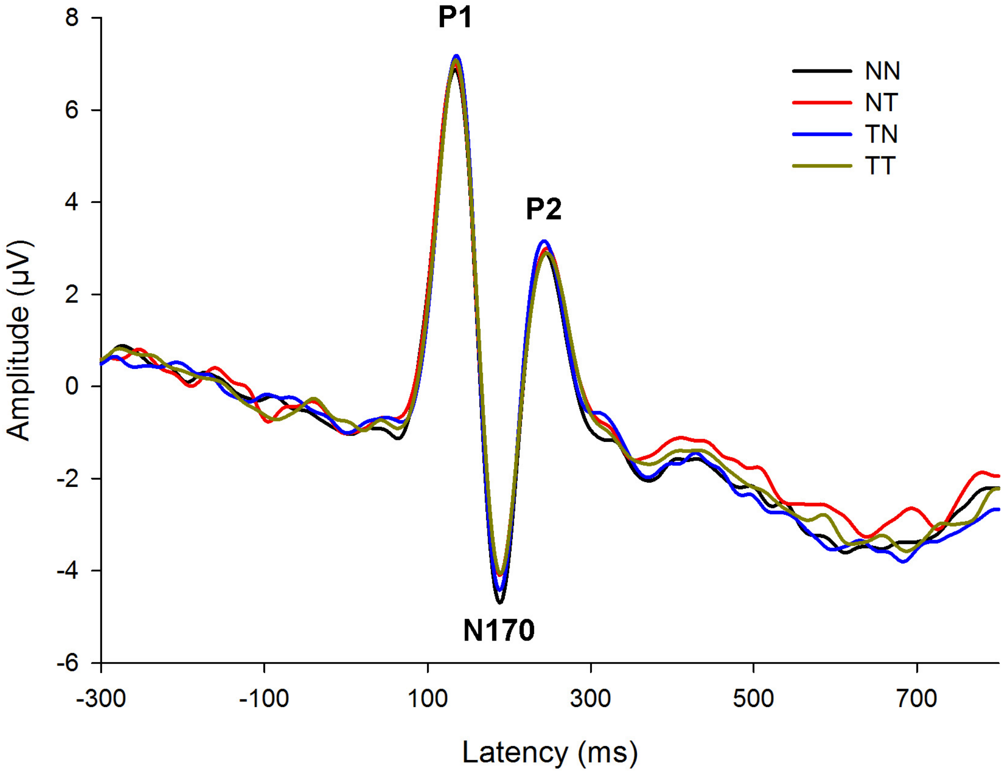


Figure 3.14. Amplitude (μV) across each inverted face condition at the PO_8 electrode.

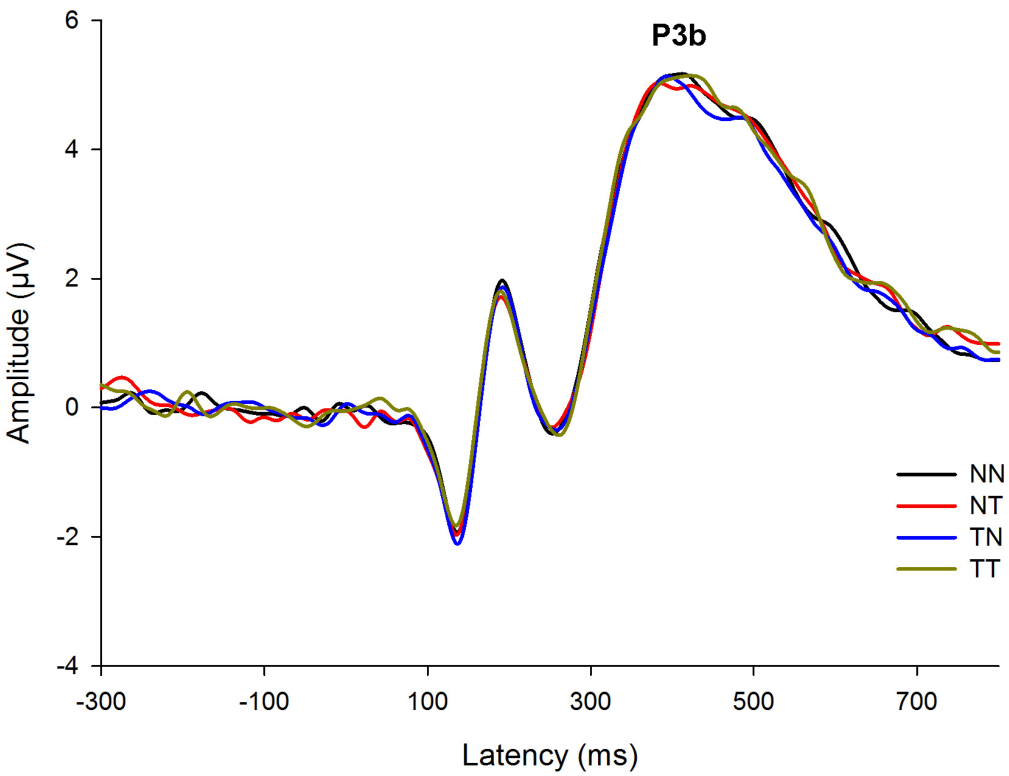


Figure 3.15. Amplitude (μV) across each inverted face condition at the CP_z electrode.

Results Summary

Hemisphere is important for the N170, P2 and P3b. With respect to the P2 and P3b components, the effect of hemisphere is consistent with a speeded peak P2 amplitude in the right hemisphere and an enhanced peak P3b amplitude at the midline. For the N170, hemisphere is important only in that there is an effect of condition on peak amplitude in the right but not the left hemisphere. Orientation influenced the amplitude of the N170, P2 and P3b. For the N170, inverted faces led to greater peak amplitudes compared to upright faces. This effect was reflected through to the P2 where inverted faces led to reduced peak amplitude relative to upright faces. At the P3b, inverted faces led to greater peak amplitudes compared to upright faces. Inverted faces also increased the latency of peak amplitude of the N170 compared to upright faces.

This study has shown: (1) an effect of Thatcherisation on faces at N170, defined by an effect of condition on amplitude in the right hemisphere; (2) an inversion effect on latency and amplitude on the N170, where the amplitude values for upright faces sit above that of inverted faces (this effect is also found at the P2 but only on amplitude); (3) a different effect of inversion on the P3b, where inverted faces elicit a greater amplitude than upright faces; and (4) no effects at the P1 component. Therefore, this study has shown an effect on face configuration on the N170 in the right hemisphere, and two distinct effects of inversion: one found with early components (N170 and P2) in the occipito-temporal region and another with the later component (P3b) in the central-parietal region.

General Discussion

It was shown in Chapter 2 that when the eyes and mouths of faces are shown upright or inverted, face perception is subject to three independent effects. These effects reflect differences in between-feature encoding and sensitivity and decisions-making for upright and inverted faces. The primary goal of Chapter 3 was to explore if ERP components exist which might reflect these three independent effects found when discriminating the orientation of the eyes and mouth in upright and inverted faces.

Participants reported the orientation of faces in which the orientation of eyes and mouths had been manipulated independently of the overall face. First, the results showed effects of face orientation on the N170, P2 and P3b components. The effects were qualitatively similar across N170 and P2 components due to right hemisphere localisation (e.g. Rossion & Jacques, 2008; Milivojevic et al.; 2003; Boutsen et al., 2006), but qualitatively different from the effect at the P3b component due to a midline localisation of largest amplitudes (e.g. Polich, 2012).

Results have also revealed that stimulus type interacted with hemisphere at the N170 alone, showing reduced amplitude with level of Thatcherisation in the right hemisphere, consistent with Boutsen et al. (2006). The mechanism of how the Thatcherised manipulation impacts face processing is unclear. Thatcherisation may represent face distortion or distinctiveness as both are shown to modulate N170. The amplitude of N170 was larger for unaltered faces judged to be atypical and unattractive compared to those judged to be typical and attractive (Halit et al., 2000). However, artificially distorted faces (with internal components altered) led to a reduction in amplitude at N170 compared to typical faces (Bentin et al., 1996). An alternative explanation is that Thatcherised faces may represent a class of emotional faces (e.g. Donnelly et al., 2011). There is evidence that some facial expressions modulate the amplitude of the N170 and that negative emotions peak later with distinctive activation (Batty & Taylor, 2003). There was no evidence for an interaction of orientation and Thatcherisation that would fit behavioural evidence of the Thatcher illusion. However, this failure to map behavioural effects of the illusion was also the case for an fMRI study of the Thatcher illusion (Rotshtein, Malach, Hadar, Graif, & Hendle, 2001).

It appears that there are three independent effects present in the ERP trace (two of orientation, and one of Thatcherisation), as well as three independent effects in the GRT analysis (violations of PI, PS and DS). The key issue is whether there is any correspondence between the effects found in the ERP trace and those found in the GRT analysis. Any presumed correspondence must be viewed as indicative rather than definitive. Nevertheless, it seems reasonable to think of perceptual effects as occurring earlier than decisional effects. For example, Philiastides and Sajda (2006) found effects at a later component (P3b) are more closely linked to recognition decisions with faces than effects at N170. As such, the effect at the P3b component is more likely to reflect a decisional rather than perceptual effect. Likewise the N170 and P2 components are more likely to reflect perceptual as opposed to decisional effects. Sources of the P3b are diffuse, but importantly, they are distinct from the 'perceptual' regions for visual stimuli.

It is possible that this kind of mapping can go further. The influence of between-feature interactions is likely to be reflected in graduated differences between types of stimulus. In contrast, the influence of orientation is likely to be reflected in a difference between upright and inverted faces. Consistent with these observations, only the N170 in the right hemisphere is graduated by stimulus type, whereas the N170, P2 and P3b are influenced by orientation. To the extent that it is possible to map between effects present in the ERP traces and the GRT analyses, it is possible that the right hemisphere localised Thatcherisation effect at N170 reflects violations of PI, the effect of orientation on the N170 and P2 reflects violations of PS, and the effect of orientation on the P3b reflects violations of DS. However, the main difficulty in mapping between these findings and those previously reported in Chapter 2 was the failure to show a condition by orientation interaction.

A previous attempt to map the implications of a GRT analyses in face perception is a study reported by Kuefner, Jacques, Alonso-Prieto and Rossion (2010). They measured ERPs in response to composite faces, due to Richler, Gauthier, Wenger and Palmeri (2008) having reported evidence for violations of DS in the composite face effect. The role of decisional processes in the composite face effect has been confirmed in Chapter 2. However, Kuefner et al. (2010) found no evidence of a decisional effect on the P3b. They did note an enhanced N170 when either the top or bottom half of faces changed across stimulus and probes (relative to same trials). On the basis of these data, Kuefner et al. (2010) argued that the composite face effect is perceptual in nature and not subject to decisional effects. However, two points are noteworthy. Firstly, Kuefner et al. (2010) used a partial design rather than the complete design required to exclude decisional influences on the composite face effect which leads to top down influences (Richler, Cheung & Gauthier, 2011b). Secondly, in Chapter 2 it has been shown that some violations of PS are evident in the composite face task (Feature Identity Task), especially when data are subject to probit analysis rather than marginal analysis. It was concluded that the composite face effect can be subject to violations of PS, consistent with the data reported by

Kuefner et al. (2010). However, the Kuefner et al. (2010) study may not have been sensitive to evidence of decisional effects. Therefore, the present study is the first to unambiguously show the possibility that ERP and GRT analyses may reveal related findings.

Conclusion

In summary, this study demonstrates the effect of Thatcherisation and separate effects of inversion in the ERP trace. As the effects in the ERP are independent, for the first time, we can see the influence of multiple levels of 'faceness' (Thatcherisation) in a manner that is independent of inversion. The two distinct effects of inversion suggest evidence for different sources of effects: one found with early components (N170 and P2) in the occipito-temporal region and another with the later component (P3b) in the central-parietal region. Also, these effects can be demonstrated as markers of configurality based on mapping between the three independent effects shown by the GRT analysis, providing converging evidence for evidence for multiple sources of configurality.

Chapter 4

The Development of Configural Face Processing and the Thatcher Illusion

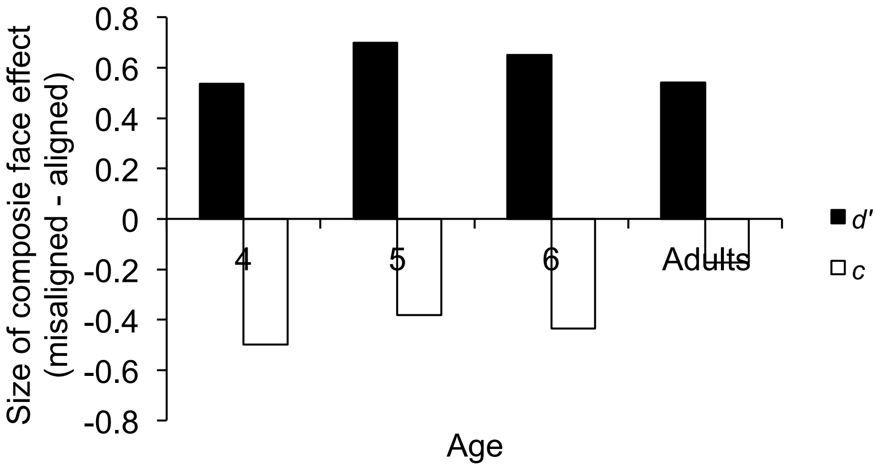
Previous chapters have shown general recognition theory (GRT), especially when underpinned by probit analysis, to be a useful tool for identifying sources of configurality in face processing. The constructs identified were also shown to correspond to effects at components in event-related potential (ERP) waveforms, at least in relation to perception of the Thatcher illusion. Chapters 2 and 3 provided converging evidence for the value of conceptualising configurality in terms of violations of perceptual independence (PI), perceptual separability (PS) and decisional separability (DS). In this chapter, I examine evidence that GRT can provide insights into face processing in a developmental study.

Despite much research, there is still considerable debate regarding how configural face processing develops across infancy, childhood, adolescence and onwards into adulthood. For many years, the consensual view of face processing development was that children younger than ten years of age process faces featurally. In contrast, children older than ten years of age process faces configurally (Carey & Diamond, 1977). This was determined through the exploration of how paraphernalia (e.g. hats, glasses, etc.) interfered with face recognition. Carey and Diamond (1977) reported that face recognition in children younger than ten years of age was markedly affected by paraphernalia, but that the face recognition of older children and adults was not. Also, the effect of inversion on face recognition was larger compared to house recognition for ten-year-olds, but not the younger children. Carey and Diamond (1977) interpreted this evidence as reflecting a processing shift. However, it is possible to provide many alternate accounts of this investigation; for example, their study might reflect the changing role of distraction on face processing with age.

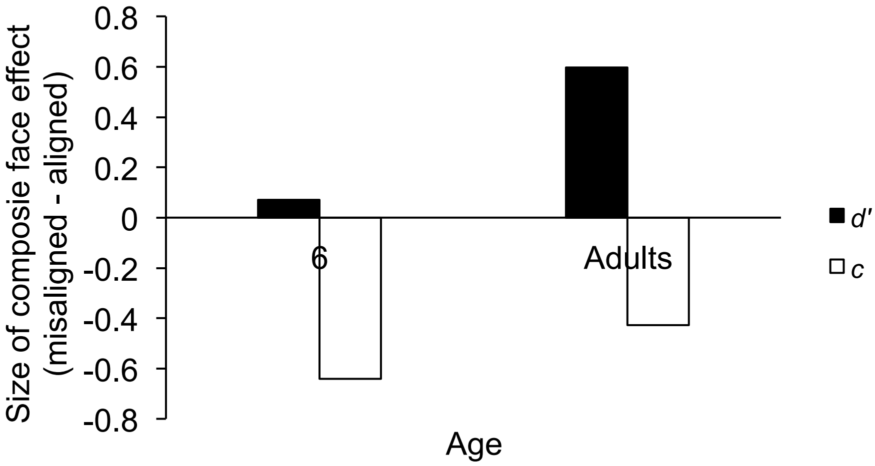
More recent studies have sought to explore face processing using more direct tests of face processing. As a result of these studies, the general consensus has shifted suggesting evidence for quantitative improvement in configural face processing with age. Using the definitions and markers of configurality discussed in Chapters 1 and 2, one group of studies have concluded configural processing is present at a young age but continues to develop; either through development of the face processing system itself (Carey & Diamond, 1994) or through general maturation (Crookes & McKone, 2009; Pellicano, Rhodes, and Peters, 2006). These studies have examined development by measuring inversion effects across faces and objects (Crookes & McKone, 2009), sensitivity to features and the spacing of features (e.g., Mondloch, Le Grand, & Maurer, 2002; Mondloch, Dobson, Parsons, & Maurer, 2004), the size of the composite face effect (e.g., Carey & Diamond, 1994; de Heering, Houthuys, & Rossion, 2007; Mondloch, Pathman, Maurer, Le Grand & Schonen, 2007), part-whole effects (e.g., Pellicano & Rhodes, 2003) and sensitivity to the Thatcher illusion (Donnelly & Hadwin, 2003).

Studies supporting the emergence of configurality propose that there is no qualitative change in face processing with development, but there may be quantitative improvement. Evidence for and against qualitative change in face processing with age has predominantly arisen from behavioural studies. However, throughout this thesis I have argued the need to improve the mapping of data to theory using more sophisticated frameworks. At present, it is unclear what is, and perhaps conversely what is not, developing with age. Rather than considering a single source of configurality and how this develops, it is possible that there is separate and independent development in PI, PS and DS. In other words, we may not have come close to understanding how the different sources of configurality in face perception develop. In this chapter, I will begin to examine this issue.

It is possible to provide some initial data to support the argument that the development of face perception is underpinned by both perceptual and decisional changes. These data come from re-analysis of findings from published studies reporting percentage accuracy data, but where the experimental designs (and reported data) allow signal detection analysis to be conducted. For example, in Experiment 2 of de Heering et al. (2007), participants were asked to report the top half of simultaneously presented composite faces as the same or different across aligned and misaligned trials. The authors reported strong composite face effects across all age groups and concluded that children as young as four process faces holisitically. When their data are reanalysed, and reported here in terms of a difference score across aligned and misaligned conditions, they show a shift in bias from age four to adulthood (Figure 4.1a). A similar re-analysis is possible from a study conducted by Mondloch et al. (2007). In this experiment, participants matched top half of composite faces across aligned and misaligned trials in a successive matching paradigm. Mondloch et al. (2007) report the magnitude of the composite face effect is the same across six-year-olds and adults. This re-analysis shows an increase in sensitivity accompanied by a decrease in bias (Figure 4.1b). Interestingly, the data from both de Heering et al. (2007) and Mondloch et al. (2007) show participants become more conservative with age.



A



B

Figure 4.1. Graphs showing sensitivity (d') and bias (c) for reanalysis of (A) de Heering et al. (2007) Experiment 2 and (B) Mondloch et al. (2007) Experiment 1.

The Thatcher illusion is an ideal paradigm to explore the development of face processing as there is no memory component that could potentially contribute to evidence for development (see Crookes & McKone, 2009). Previous studies examining a developmental sample have shown children as young as six years old are sensitive to the Thatcher illusion (Lewis, 2003; Donnelly & Hadwin, 2003). Lewis (2003) found no effect of age in the grotesque to normal switch in the Thatcher illusion when participants were asked to report the angle of change. All participants (aged 6-75 years) were sensitive to the illusion. Donnelly and Hadwin (2003) explored the Thatcher illusion with children (aged 6-12 years) and adults. All participants were sensitive to the illusion using typical versus Thatcherised faces. When processing demands were increased by using typical and Thatcherised 'Mooney' faces, sensitivity to the task increased with age and was absent for the six-year-olds.

Children with autism are able to perceive the Thatcher illusion (Rouse, Donnelly, Hadwin & Brown, 2004) and Bentin and Bhatt (2004) suggested 6-month-old babies are sensitive to the Thatcher illusion although they used an indirect measure of assessing sensitivity using habituation-novelty preference unlike previous studies using direct measures of the illusion (Lewis, 2003; Donnelly & Hadwin, 2003). Therefore, evidence suggests the illusion is reliably perceived by young children.

Experiment 4.1

Chapter 2 has outlined the value of the GRT approach with the Thatcher illusion to reveal sources of configurality. Thus, in order to explore the development of different sources of configurality, Chapter 4 reports a developmental study of the Thatcher illusion within a GRT paradigm. As large numbers of trials are required to perform the GRT analyses, data will be aggregated from groups of children of similar age as it was not possible to collect the volume of data required from individual children. Aggregate analysis for GRT has been reported previously (e.g. Wenger & Ingvalson, 2002; Richler, Gauthier, Wenger, & Palmeri, 2008), but not with children.

Aims

(1) To examine whether the task is appropriate for children aged eight years old and to assess whether children are able to respond to both features without a compensatory strategy (focusing on just one feature).

(2) To determine whether there is a change in performance with age.

(3) To examine how sources of perceptual and decisional violation develop with age.

Method

Participants

Participants were recruited from three separate age brackets. Twelve children were recruited for age group one (age range 8-9, M = 8.67, SD = 0.49), six were female. Ten children were recruited for age group two (age range 10-11, M = 10.70, SD = 0.48), eight were female. Children were recruited at their school via opt-in consent from parents or via opportunity sample, again with opt-in consent from parents. Ten adults (age range 22-34, M = 27.40, SD = 3.92) were recruited via opportunity sampling at the University of Southampton, six were female. All participants had normal or corrected to normal vision.

Design

The task required participants to judge the status of two features (the eyes and mouth) across two levels (normal versus inverted orientation). Together, these two dimensions, each with two levels, created four stimulus conditions. There were two versions of the task, one with upright faces and the other with inverted faces. These were completed as separate blocks.

Apparatus and Stimuli

The tasks were built in Experiment Builder (Version 1.5.201). The stimuli were the same as those used in Experiment 3.1 (ERP Study, see Chapter 3) and appeared centrally at a size of 4.70 cm by 3.30 cm on the screen. All stimuli were presented against a black background on a laptop computer with a screen size of

12.20 cm by 30.50 cm, resolution of 1024 by 768 and refresh rate of 60 Hz. Reponses were made via a mouse button press on the laptop touch pad. All text was presented in white. Prompts when responses were required consisted of ‘normal’ and ‘rotated’ being displayed on the side of the screen corresponding to the button along with a prompt for which question to respond to first indicated by the location of the words on the screen (either closer to the top or closer to the bottom of the screen). The approximate viewing distance was 60 cm.

Procedure

Participants completed upright and inverted versions of the task within the same session. There were 40 trials in each orientation task, 10 for each of the four conditions. All trials were randomised within each task, with breaks to create four blocks of 10 trials. Button press, feature to respond to first and order of orientation version of the task were randomised across participants. For most children, two participants completed the task in the same room at the same time; thus, both participants had the same counterbalance version as instructions were given to both children at the same time. All adults completed the task one at a time.

Participants were asked to maintain their head in an upright position and this was checked by the experimenter. Each task began with a 500 ms fixation cross presented centrally on the screen. The stimulus was presented for 1000 ms followed by the first prompt to respond. Once the first response was made the second prompt to respond was displayed. Once the second response had been made a noise mask was presented for 100 ms before the next trail began. No feedback was given for responses.

Results

All responses were included regardless of speed of response. Firstly, group based analysis was used to examine patterns of performance within the task based on age and task conditions. Then, aggregate analysis was used to examine evidence for configurality via GRT.

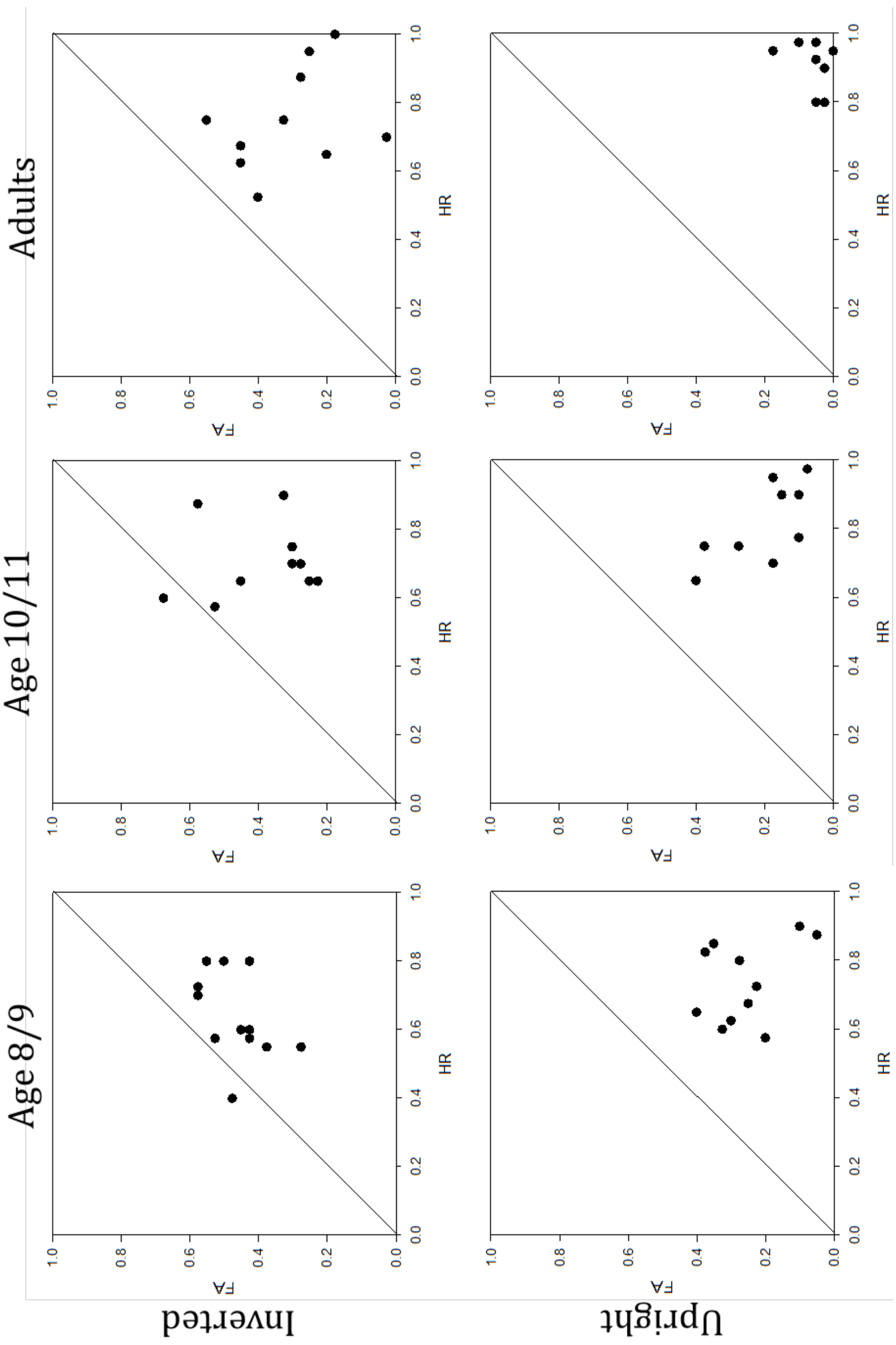


Figure 4.2. Graphs showing hit rate (HR) and false alarm rate (FA) for responses to eyes. Graphs panelled by age group and orientation condition.

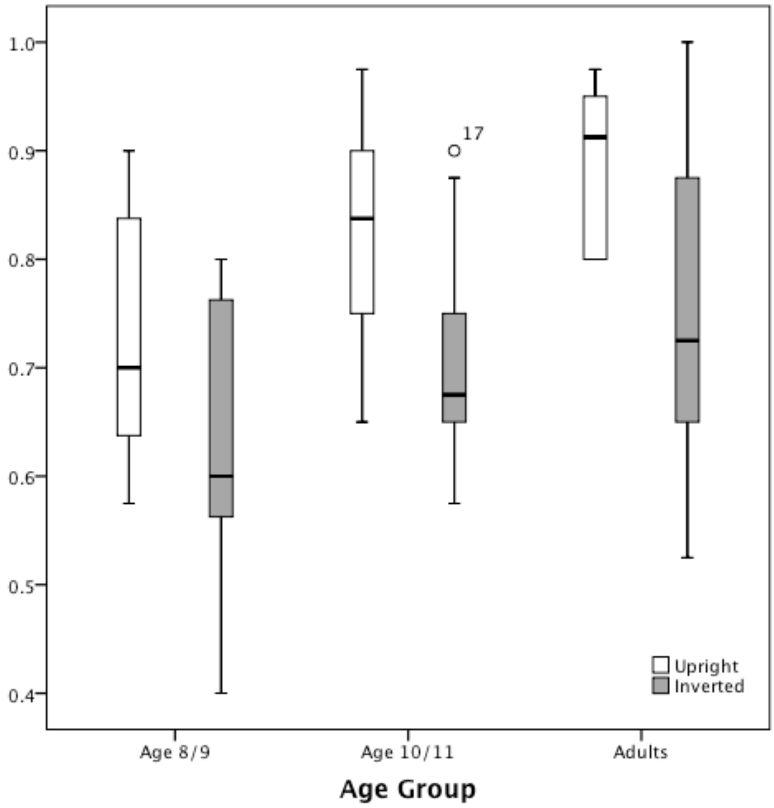


Figure 4.3. Boxplot showing the range of hit rate (HR) values across the participants for the three age groups.

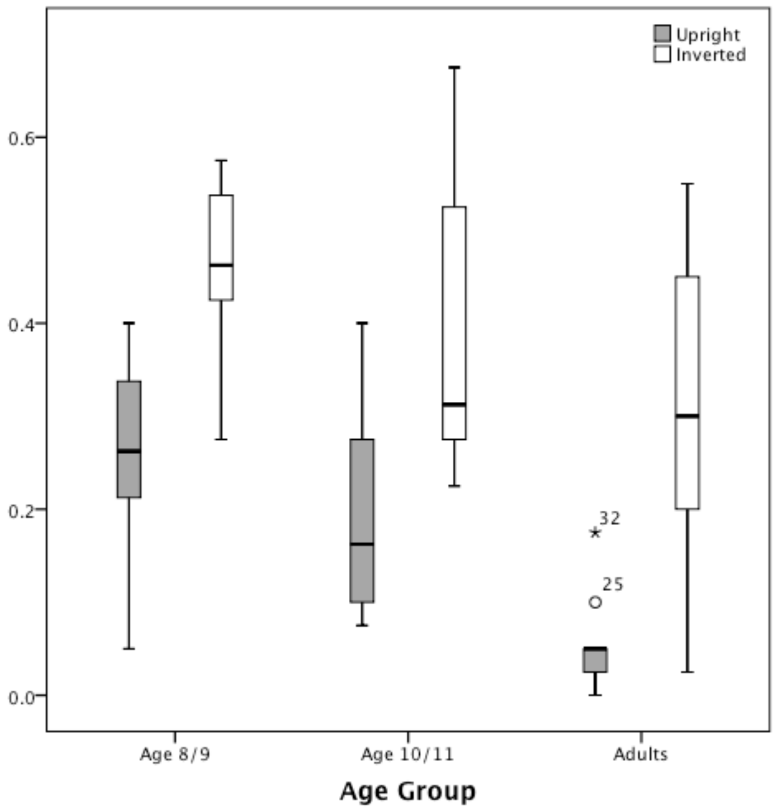


Figure 4.4. Boxplot showing the range of false alarm (FA) values across the participants for the three age groups.

Group-based Results

Across three age groups, it can be seen that performance improves with age (Figure 4.2). Hit rate (HR) improves and false alarm (FA) rate reduces, with more points at the lower right hand corners of the graphs and further from the line representing chance performance. Also, at all ages, performance in the upright conditions is better than for the inverted condition. For the children, there is more variation in performance in the upright condition, whereas the adults group performed uniformly on the task. For the inverted condition, performance is varied within each age group.

When comparing the spread of HR (Figure 4.3) and FA (Figure 4.4) values it is clear there is a large variation in HR and FA across the groups and the spread appears to reduce with age for the upright condition only.

Sensitivity (d') and bias (c) values for the overall task were calculated for each participant using the HR and FA data and the 2AFC signal detection theory (SDT) paradigm. For a HR of 1, the correction 1- $\frac{1}{2N}$ was applied (Macmillan & Creelman, 2005). The d' and c data were analysed across two mixed factorial ANOVAs with group (age 8/9, age 10/11 and adults) as a between subjects factor and orientation (upright and inverted) as a within subjects factor. For d', the main effect of orientation was significant (F(1,2) = 138.71, p < 0.001, η_p^2 = 0.04). The d' for the upright condition was greater (M = 1.58, SE = 0.06) than the inverted condition (M = 0.71, SE = 0.06). The effect of age group was significant (F(1,2) = 28.26, p < 0.001, η_p^2= 0.31). Sensitivity increased with age for the 8/9 year olds (M = 0.73, SE = 0.08), the 10/11 year olds (M = 1.10, SE = 0.09) and adults (M = 1.60, SE = 0.09), with a significant difference in d' between the 8/9 year olds and adults (p < 0.001) and the 10/11 year olds and adults (p < 0.001). The interaction of age group and orientation was non-significant (F(1,125) = 2.64, p = 0.075).

For c, the main effect of orientation was significant (F(1,2) = 14.82, p < 0.001, η_p^2 = 0.11). Bias for the upright condition was more conservative (M = 0.05, SE = 0.04) than for the inverted condition (M = -0.14, SE = 0.05) (more likely to say the feature is normal when inverted). The effect of age group did not reach significance (F(1,2) =

0.59, p = 0.557). The interaction of age group and orientation did not reach significance (F(2,125) = 0.45, p = 0.640, see Figure 4.5).

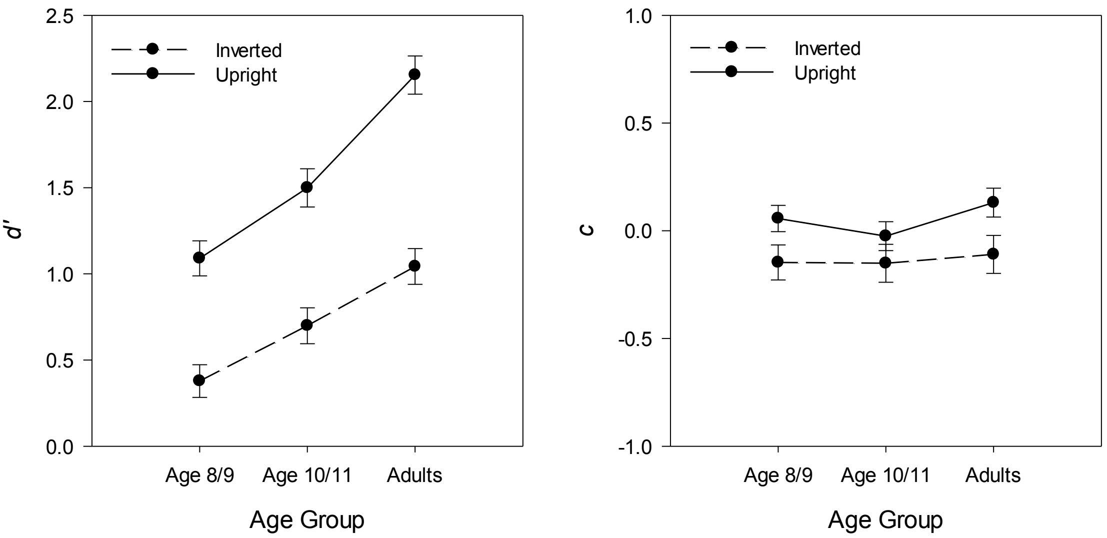


Figure 4.5. Plot showing d' (left) and c (right) by age group and orientation in the Thatcher illusion task.

Aggregate Results

For each participant, a confusion matrix of responses was created for each orientation version of the task. The confusion matrices were aggregated across all participants in the age range. This created six matrices, one for each orientation condition for each of the three age groups. Each matrix was based on at least 100 trials of each stimulus type in each orientation condition of the task. If there was a zero for any cell in the matrix, then one was added to all values in the matrix to allow computation of multidimensional signal detection analysis (MSDA) measures. Marginal sensitivity (*d'*) and bias (*c*) were calculated using the equations outlined in Chapter 2 from the aggregated data across participants. The values for *d'* and *c* for each marginal condition are presented in Figure 4.6 and Figure 4.7.

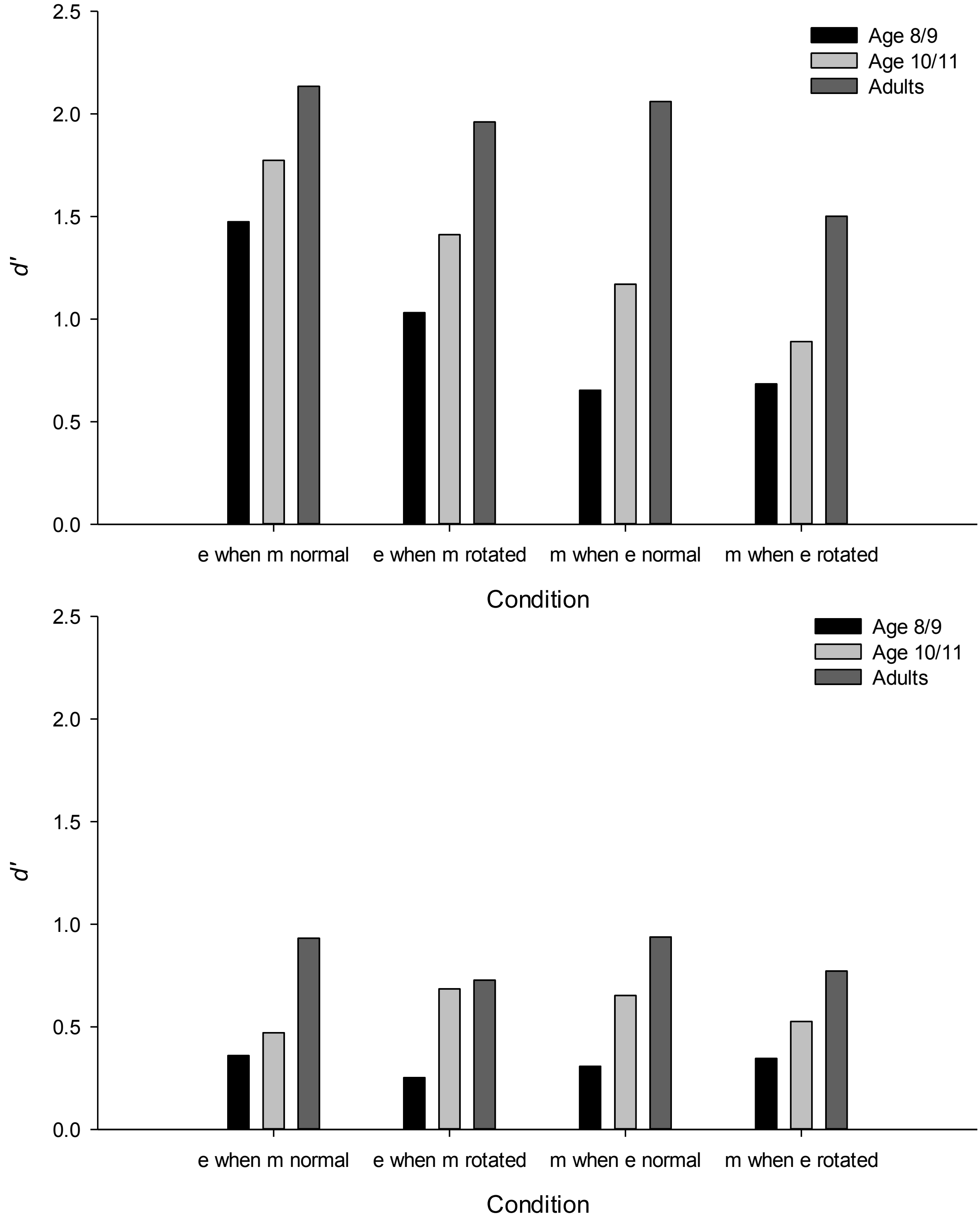


Figure 4.6. Graphs showing the sensitivity d' for upright (top) and inverted (bottom) conditions in the Thatcher illusion task. For the marginal conditions, e = eye and m = mouth.

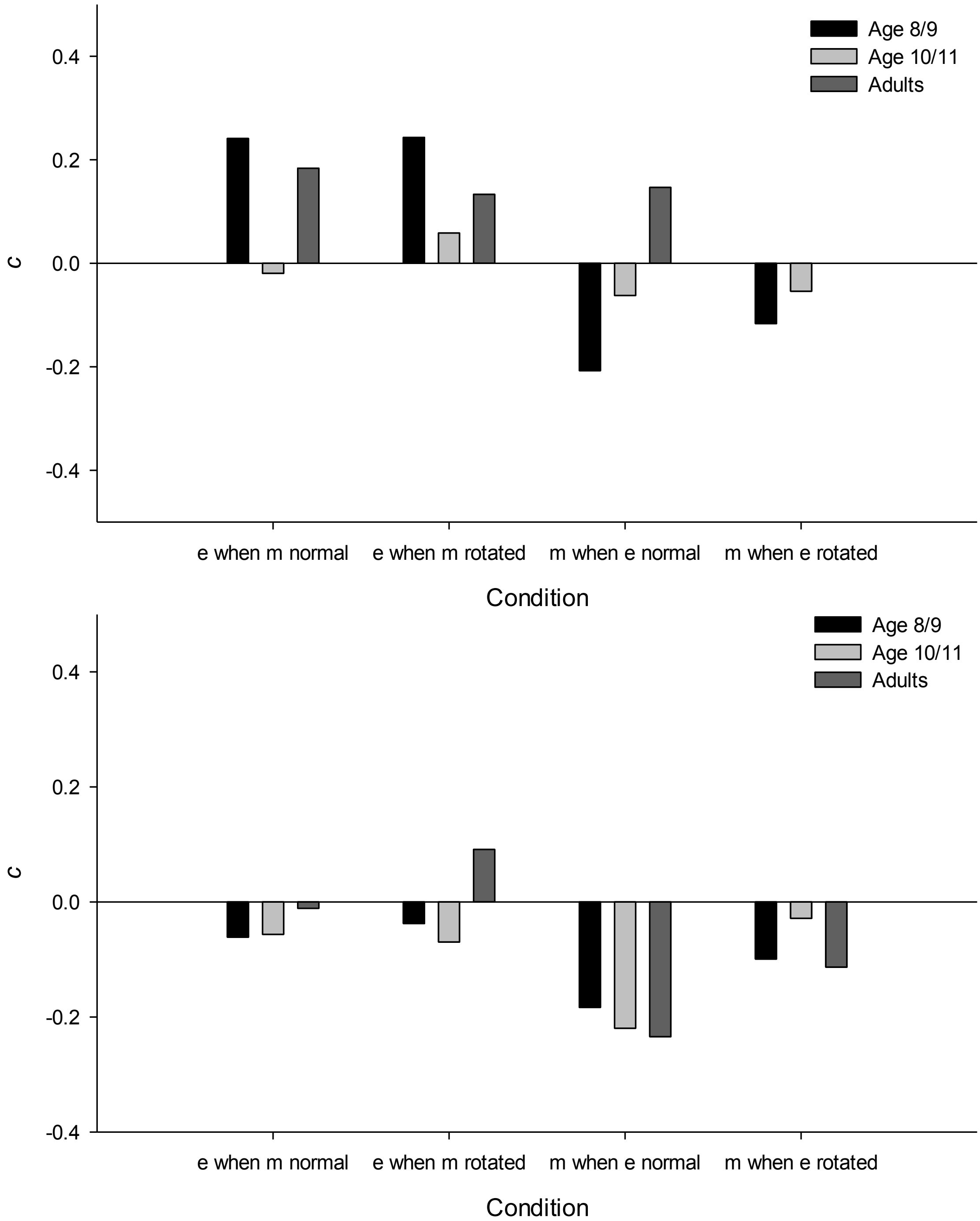


Figure 4.7. Graphs showing the bias (c) for upright (top) and inverted (bottom) conditions in the Thatcher illusion task. For the marginal conditions, e = eye and m = mouth.

With only one value per age group, evaluating a statistical difference across age for the marginal sensitivity and bias data is not possible. However, the descriptive results reveal marginal sensitivity improves with age for both the upright and inverted conditions. In the upright condition, children appear to be more sensitive to the eyes than the mouth. In particular, children appear to be most sensitive in detecting the eyes when the face is upright and the mouth is in the 'normal' orientation. There appears to be little bias across age groups in the inverted condition, with a slight bias to response liberally to label the feature as 'normal'. Conversely, in the marginal condition of 'mouths when eyes are normal' where all age groups show a liberal bias to responding mouths are 'normal'. For the upright condition, the youngest age group of 8/9 year olds show the largest bias values across marginal conditions. Predominantly, the younger children are biased to say the eyes are 'rotated' and the mouths are 'normal' in upright faces which can be seen by the bias values swapping from positive to negative values with feature. At the superficial level, it appears bias is reducing with age for the upright condition.

Table 4.1

Summary of Marginal and Probit Violations for each Age Group in each Orientation Condition

			Inferences							
	Orientation	Participant	PI11	PI21	PI12	PI22	PSh	PSv	DSh	DSv
Marginal	Upright	8/9 Year Olds	0	0	0	0	0	1	0	0
		10/11 Year Olds	0	0	0	0	0	0	0	0
		Adults	0	0	0	0	1	0	?	0
	Inverted	8/9 Year Olds	0	0	0	0	0	0	0	0
		10/11 Year Olds	0	0	0	0	0	0	0	0
		Adults	0	0	0	0	0	0	0	0
Probit	Upright	8/9 Year Olds	0	1	0	0	0	0	0	0
		10/11 Year Olds	0	0	0	0	0	0	0	0
		Adults	0	0	0	0	0	0	0	0
	Inverted	8/9 Year Olds	0	1	0	0	0	0	0	0
		10/11 Year Olds	0	0	0	0	0	0	0	0
		Adults	0	0	0	0	0	0	0	0

Note. Violations of equalities and measures are noted as '1', the measure holding is noted as '0' and '?' indicates an uncertain inference. For the numbers after the measures of PI, these refer to the particular distributions of which there are four. For the letters after the measures, 'h' denotes the measure in the horizontal (for the mouth across levels of the eyes), 'v' denoted the measure in the vertical (for the eyes across levels of the mouth). Violations of inferences are highlighted in grey.

Probit analyses[5] were completed using the methods described in Chapter 2 and the results are presented in Table 4.1. The only significant violation was for 8/9 year olds where there was a violation of PI in both upright and inverted faces, driven in both cases by negative correlation in the eyes-rotated- mouth-normal condition. There were no violations found for the 10/11 year olds or the adults in either the upright or inverted condition. The correlations observed for each distribution for each age group are presented in Table 4.2.

The two violations of PI for the 8/9 year olds were driven by negative correlations. The correlation values were examined for inversion effects across a particular distribution as described in Chapter 2. Results revealed there was no inversion effect for the 8/9 year olds across the X_2Y_1 distribution which showed violations of PI across both orientation versions of the task. However, the rho value for the adult X_2Y_1 upright condition did not fall within the 95% confidence interval of the rho for the X_2Y_1 distribution in the inverted condition, so an inversion effects was observed despite no violations of PI in this distribution. An inversion effect was not observed for any of the other distributions across any of the age groups (Table 4.2). This effect is suggestive of an emerging effect of orientation on the between–feature encoding of inverted eyes and upright mouths. Note that it is this same comparison that underpinned the violations of PI reported in Chapter 2.

[5] For the marginal analyses, no violations were observed for the inverted version of the task for any of the three age groups. For the upright version of the task, violations of PS were observed for the 8/9 year olds and adults only, although the origin of the violation differed.

Table 4.2

Correlation Value for each Distribution in each Orientation Condition for each of the Age Groups and Inversion Effects for Values in each Distribution using 95% Confidence Intervals

		X_1Y_1	X_2Y_1		X_1Y_2	X_2Y_2
8/9 Year-Olds	Upright	**-0.05**	**-0.48**		**-0.21**	**-0.25**
		-0.49	-0.92		-0.56	-0.69
		0.39	-0.04		0.15	0.19
	Inverted	**-0.06**	**-0.38**		**-0.24**	**-0.19**
		-0.44	-0.71		-0.59	-0.54
		0.32	-0.04		0.11	0.16
10/11 Year-Olds	Upright	**0.23**	**-0.21**		**-0.22**	**0.00**
		-0.35	-0.75		-0.69	-0.52
		0.80	0.32		0.26	0.52
	Inverted	**-0.29**	**-0.17**		**-0.14**	**0.15**
		-0.80	-0.62		-0.55	-0.27
		0.23	0.28		0.26	0.58
Adults	Upright	**-0.05**	**-0.48**		**-0.17**	**0.04**
		-0.76	-1.11		-0.95	-0.70
		0.65	0.14	*	0.60	0.79
	Inverted	**-0.05**	**0.18**		**-0.12**	**-0.14**
		-0.50	-0.22		-0.54	-0.55
		0.41	0.59		0.29	0.27

Note. Distributions showing violations of PI are highlighted in grey. Bold numbers represent the mean rho values for each distribution and the values below the mean represent the confidence interval. * indicates significant difference between rho values if the mean of one distribution is not within the confidence interval of the other mean rho values.

Discussion

The group based analysis suggested sensitivity improved with age. This type of evidence has previously been used to suggest a component of face processing continues to develop with age (e.g. Mondloch et al., 2002; Donnelly & Hadwin, 2003). Inversion effects were found for sensitivity to the illusion, for each age group. Again, this has been taken previously to indicate configural processing in all age groups tested (e.g. Pellicano et al., 2006). Therefore, the basic sensitivity data confirm that young children are sensitive to the Thatcher illusion. Thus, all prior experimental data from similar tasks are confirmed.

If one assumes that the inverted condition represents baseline in this task, then the data also show that there are different baselines for the various age groups. Varying baselines compromise the comparison of evidence for configurality across ages (as critiqued by Crookes & McKone, 2009). GRT can overcome this as the nature is determined from the group's data independent of the other age groups and also allows conclusion about the effects observed in the ANOVAs. In this task, GRT was used to explore whether the change in face processing with age is underpinned by qualitative change or quantitative improvement across multiple sources of configurality. Problematically, the basic GRT analysis using the probit model did not appear consistent with the findings of Chapter 2. In terms of replicating results from the Feature Orientation Task in Chapter 2, no violations of DS were observed when expected for the inverted condition. Using the probit model, violations of PS were not observed when they were expected in the upright condition. In addition, violations of PI were not observed for the adults when they were previously found for both orientations. The reasons for these different findings across Chapters 2 and 4 are discussed shortly.

With respect to the current experiment, there were no violations of PS or DS and only the 8/9 year olds showed violations of PI. This violation of PI was present in the upright and inverted conditions. Violation of PI for the 8/9 year olds in the eyes-rotated-mouth normal condition was consistent with previous findings. It seems likely that the current experiment is underpowered to reveal the violations that might emerge. Two facts point to this view. Firstly, the correlation across inverted eyes and upright mouths, when faces were upright, was numerically the same in adults as in 8/9 year old children. Secondly, when comparing the rho values in the X_2Y_1 distribution across upright and inverted conditions, there was a significant inversion effect for adults, but not children. The upright condition showed a large negative correlation compared to a small positive rho value in the inverted condition. This effect for adults is similar to that reported previously in Chapter 2. Therefore, there is some indication of consistency to my previous finding with adults that is not apparent from the violations alone.

This study examined whether the GRT framework would be appropriate to use with children if each participant only contributed a small amount of data to the overall aggregate matrix. Aggregate analysis has been used previously in adults for GRT analysis (e.g. Wenger & Ingvalson, 2002, 2003; Richler, Gauthier, Wenger, & Palmeri, 2008). However, each participant in these previous studies contributed a larger number of trials than was possible in the present study. The present experiment should be taken as indicative of the fact that it is possible to run a full factorial experiment with young children. Nevertheless, the fact that the underlying violations did not robustly replicate those found in Chapter 2, suggests more participants are required than were tested here.

The present experiment differed in a number of ways from that reported in Chapter 2, such that the task demands were altered. In the present experiment, the stimuli were better matched across conditions in terms of low level properties, the exposure duration was longer and there was no backwards masking. Also, a speeded response was not required. It may be that the differences between the results of Chapter 2 and the present experiment reflect these differences rather than reduced statistical power.

Conclusion

The study is underpowered and, therefore, conclusions about quantitative versus qualitative change in processing with age for each of the types of configurality identified in Chapter 2 are premature. However, there is sufficient power in the data to replicate all previous findings, when the data are considered at the behavioural level in terms of sensitivity. Furthermore, the probit GRT analysis was still able overcome the issue of different baselines to reveal evidence of between-feature configural relations in upright faces that may change with age. More data is required to properly explore all types of violations, including those decisional effects suggested by the reanalysis of previous experimental data. Nevertheless, this study has shown the GRT framework is a useful tool for examining the development of sources of configurality.

Chapter 5

Prosopagnosia and the Thatcher Illusion

In Chapters 2 and 3, I have explored sources of configurality in face processing, with a specific emphasis on the Thatcher illusion. Further to this, Chapter 4 showed how this approach has value in studying the development of face processing. In this final empirical chapter, I finish by exploring whether the general recognition theory (GRT) framework can be applied to understand the perception of the Thatcher illusion in a case of acquired apperceptive prosopagnosia (Bodamer, 1947). Somewhat problematically, there are very few published reports of how the Thatcher illusion is perceived in apperceptive prosopagnosia (Boutsen & Humphreys, 2002; Carbon, Grüter, Weber, & Lueschow, 2007). To provide a sound basis from which to explore the value of GRT to the understanding of prosopagnosia, the present chapter begins with an exploration of the perception of the Thatcher illusion in apperceptive prosopagnosia. Data from Chapter 5 has been published in Neuropsychologia (Mestry, Donnelly, Menneer, & McCarthy, 2012).[6]

Prosopagnosia is a condition in which individuals have selective difficulties in recognising faces. Deficits can be acquired (e.g. through brain injury) or can be developmental (also referred to as congenital). Types of prosopagnosia are distinguished by the deficits in behaviour; thus, the disorder is often known by the outcome rather than the cause. Associative prosopagnosics have normal vision and perception, but a limited understanding of what a face represents. Apperceptive prosopagnosia is more severe as individuals have a deficit in perception of a face. These individuals have more difficulty copying and recognising faces as well as with face-matching tasks (Goldsmith & Liu, 2001).

Studying an individual with a deficit in configural processing can be valuable to the understanding of the nature of configural processing in typical individuals, but

[6] The Categorisation and Discrimination tasks in Experiment 5.2 were included in my undergraduate dissertation (Mestry, 2008). Since that assessment, additional data has been collected from PHD to create triple the amount. The results reported here use signal detection analysis instead of percentage accuracy as reported previously. Thus, the same data has not been submitted for assessment twice.

the issue is complicated by the widely differing patterns of impairment, with few ‘pure’ cases. Previous studies with apperceptive prosopagnosics have attempted to identify the nature of impairments using behavioural (e.g. Farah, Wilson, Drain, & Tanaka, 1995) and neurophysiological (e.g. Eimer & McCarthy, 1999) evidence. Past results have shown intact emotion processing abilities for individuals with apperceptive prosopagnosia (Duchaine, Parker, & Nakayama, 2003), the positive influence of facial emotion expression on face identification (de Gelder, Frissen, Barton, & Hadjikhani, 2003; Peelen, Lucas, Mayer, & Vuilleumier, 2009) and the difficulty recognising negative emotions such as anger and disgust, but not happiness (Stephan, Breen & Caine, 2006). Emotion perception in prosopagnosia has been frequently explored, but there is little agreement of the impact of prosopagnosia on emotion perception.

In line with the previous studies, the Thatcher illusion will be used as a tool to explore deficits for an individual with apperceptive prosopagnosia (PHD). Donnelly et al. (2011) have shown the illusion involves activation of emotional cortices. Therefore, there is a need to explore the emotion processing capabilities in light of an illusion which is commonly thought to be driven by configural processing alone. Specifically, the processing of negative emotions will be examined due to the grotesque appearance of the Thatcherised face.

Discriminating Thatcherised from Typical Faces in a Case of Prosopagnosia

The Thatcher illusion (Thompson, 1980) refers to the change in perception of ‘Thatcherised’ faces when they are rotated from upright to inverted orientations. Thatcherised faces are made by inverting the eyes and mouth. These faces appear grotesque when upright but when inverted the grotesqueness disappears and faces appear more typical. There have been various accounts put forward to explain the Thatcher illusion. These accounts range from expression analysis (Valentine, 1988), to the phenomenological experience of conflicting reference frames for faces and facial features (Parks, Coss, & Coss, 1985). More recently, and following Bartlett and Searcy (1993), the Thatcher illusion is thought to reveal the orientation specific nature of configural face processing. Perception of grotesqueness in the upright face is attributed to the perception of unusual configural relationships between the features, whilst perception of the typical inverted face relies on feature based processing (Stürzel & Spillmann, 2000). Furthermore, the automaticity with which grotesqueness is experienced makes it a useful test of the presence of configural face processing in atypical populations (Rouse, Donnelly, Hadwin and Brown, 2004) including acquired (Boutsen & Humphreys, 2002) and congenital prosopagnosia (Carbon et al., 2007).

There are two (related) ways in which configural processing (defined here as the encoding of between-feature spatial relationships) might lead to the perception of the Thatcher illusion. Firstly, by a poor match between representations of Thatcherised and prototypical faces. Secondly, by creating local difficulties in configural processing for inverted eyes and mouths which are in an unusual orientation relative to otherwise upright faces. By the first account, grotesqueness results from the fact that individuals rate average and not distinctive faces as attractive (Rhodes & Tremewan, 1996). By the second account, grotesqueness results from low processing fluency associated with processing Thatcherised faces as faces (Reber, Winkielleman, & Schwartz, 1998). Thatcherised faces are poor examples of the face category and lead to a processing difficulty that is experienced as grotesqueness rather than slow processing.

The involvement of emotional coding, in addition to configural processing, in the Thatcher illusion is manifest in recent neuroimaging studies. Specifically, areas known to be involved in social and emotional processing are also involved in the perception of both single Thatcherised faces (Rothstein, Malach, Hadar, Graif, & Hendler, 2001), and when discriminating Thatcherised from typical faces (Donnelly et al., 2011). Therefore, despite being thought of as an illusion demonstrating configural processing in faces, and Thatcherised faces not representing standard emotional faces, the phenomenology of the Thatcher illusion depends on the response of diffuse socio-emotional cortices to Thatcherised faces. The consequence is that any use of the Thatcher illusion as a marker of configural processing should be accompanied by evidence of broadly intact emotional processing. Otherwise any failure in perception of the Thatcher illusion might result from deficits in socio-emotional processing.

It is important to understand what is meant by broadly intact emotional processing in the context of a deficit in configural processing. The issue is complicated by the fact that configural processing contributes to the perception of some facial emotions. For example, Calder Young, Keane and Dean (2000) measured response times to aligned and misaligned composite faces where the face composites are formed from the same or different emotions. By determining which emotions were responded to more quickly when the top and bottom halves of faces were aligned relative to misaligned, Calder et al. (2000) were able to determine which emotions have their perception facilitated by configural processing. Table 1 of Calder et al. (2000) indicates the perception of anger, fear and sadness from whole faces cannot be predicted from that of part faces. In contrast, the perception of disgust, happiness and surprise from whole faces can be predicted from that of part faces. This means that the detection of anger, fear and sadness is improved by computing configural information from whole faces.

The role of configurality in emotion processing was also explored by McKelvie (1995) who compared categorisation accuracy to emotional faces shown in upright and inverted orientations. Inversion led to less accurate categorisation of sadness, fear, anger and disgust compared to when upright. Happiness, neutrality and surprise were unaffected by orientation. McKelvie (1995) suggested emotional expressions rely on configurality to different degrees, similar to Calder et al. (2000). Together these studies show configurality was important in the perception of anger, fear and sadness. Only in respect of the perception of disgust did the two studies differ.

The next question asks what impact would the absence of configural processing have on emotion perception in general and the Thatcher illusion in particular? I propose that an absence of configural processing should be manifest in (1) a specific pattern of modest deficit in the recognition of facial emotions that partially rely on configural processing and (2) an inability to perceive Thatcherised faces. In this study, a series of experiments conducted on a brain-damaged patient with prosopagnosia are reported. PHD has been reported previously in an ERP study comparing unfamiliar faces and houses and does not generate an N170 component in response to faces (Eimer & McCarthy, 1999). If a failure to generate a face effect at N170 is linked to a failure of configural processing in faces through an absence of face categorisation (e.g. Eimer, 2000b), two related findings can be predicted. Firstly, PHD will be able to categorise facial emotion but only when this can be achieved from features. Secondly, PHD will be unable to categorise Thatcherised faces from typical faces.

Experiments 5.1 and 5.2

Exploration starts by testing PHD's ability to categorise facial emotions and to discriminate between faces exhibiting different levels of emotion. It could be hypothesised that PHD's categorisation of emotional faces will be impaired in those conditions that rely on configural processing. Nevertheless, it is also predicted that PHD will have intact categorisation and discrimination of emotions and emotional intensity when this can be achieved through featural analysis. Therefore, the goal of Experiment 5.1 was to determine whether PHD is able to perform categorisation and discrimination of faces with emotional valence in at least some conditions. In Experiment 5.2, PHD's ability to categorise and discriminate Thatcherised from typical faces is explored. The ability of PHD to perform in these tasks is compared with that of controls.

Aims

(1) To examine the facial emotion processing ability of PHD compared to controls.
(2) To examine the role of configurality in emotion processing.
(3) To examine whether PHD is sensitive to the Thatcher illusion.

Experiment 5.1: Emotion Perception Studies

In Experiment 5.1, the sensitivity of PHD and age-matched controls participants to categorising emotions was measured. While PHD's overall ability to categorise emotions was of interested, so was his ability to do so when only two emotions are possible. In this case, the task demands of categorisation are the same as when categorising Thatcherised from typical faces. The clinical testing reported in this method section was conducted by Professor Rosaleen McCarthy has been reported in Mestry, Donnelly et al. (2012).

Method

Participants

An individual with prosopagnosia (PHD) volunteered to participate in studies regarding his deficit. PHD is a left-handed male, who was aged 48 to 51 over the course of the current experiments. He sustained a closed head injury as a result of a road traffic accident at the age of 17. Structural MRI in 2005 (Figure 5.1) showed a unilateral lesion in the ventral temporal lobe in the region of the fusiform gyrus on the left with no other macroscopic areas of damage. PHD suffers significant cognitive deficits including apperceptive prosopagnosia and some category specific visual agnosia, especially for the living things domain (animals and fruit/vegetables). He has persistent difficulties recognising people from their faces without context or other supporting information. PHD has a mild deuteranomaly and corrected-to-normal visual acuity with eye-glasses and his visual fields are full.

See Figure 1 in Mestry, Donnelly, Menneer and McCarthy (2012) due to issues of copyright.

Figure 5.1. Structural MRI taken from PHD showing a focal area of injury in the inferior temporal lobe of the left hemisphere in the region of the fusiform gyrus.

PHD's most recent cognitive assessment showed him to be functioning at an average level on most subtests of the WAIS-III (Wechsler, 1997, Table 5.1). On the Visual Object and Spatial Perception Test battery (VOSP, Warrington & James, 1991) he scored within the normal range on Screening, Fragmented Letters, Object Decision, Dot Counting, Position Discrimination, Cube Analysis and Number Location. On the Warrington Recognition Memory Tests for Words and Faces (Warrington, 1984), PHD's recognition memory for words was above chance, but within the clinical range for his age. His recognition memory for faces was at chance, but he was within the normal range on the Camden test of memory for Topographic Scenes (Warrington, 1996, Table 5.1). When confronted with portraits of contemporary famous people, PHD was only able to identify the Queen and President Obama. He indicated some familiarity with some of the faces (e.g. asking 'is he an entertainer?' for Bruce Forsyth). If the most lenient criterion for recognition is adopted, PHD scored 6/22 correct: a score below the poorest achieved by patients attending a clinic for people with moderate levels of Alzheimer's disease (McCarthy, personal data).

Table 5.1

Performance of PHD across a range of Cognitive Assessments

Task	Score
WAIS III[a]	
Verbal Scale	
Vocabulary	9
Similarities	9
Arithmetic	5
Digit Span	10
Performance Scale	
Picture Completion	6
Digit Symbol Coding	6
Block Design	9
Matrix Reasoning	10
Camden Memory Tests[b]	
Faces	12/25 (chance)
Words	21/25 (2nd percentile)
Scenes	20/30 (7-9th percentile)
Paired Associates	17/48 (<1st percentile)
Graded Naming Test[c]	8/30 (< 1st percentile)

[a]Scores represent age-related scores, *M* = 10, *SD* = 3, Wechsler (1997); [b] Warrington, (1996); [c]McKenna and Warrington (1983).

The Benton Face Recognition Test (Benton, Sivan, Hamsher, Varney, & Spreen, 1994) was used to assess the status of face perception. PHD was able find a single face in the array of six that is identical to the target, scoring 6/6 on this first part. The remaining 16 items require matching for facial identity across viewpoints: the participant must find three examples of the target face amongst an array of six items. PHD only succeeded on 30/48 choices, scoring within the severely impaired range for the test as a whole. PHD’s score deteriorated further when the task was presented in an inverted orientation (28). He tried to solve the puzzle of matching the inverted faces by using an overt feature-naming strategy ('have they got the same eyebrows?' or 'I can tell by the cheeks').

PHD's ability to extract emotion cues from faces was initially evaluated clinically using a paper based version of the Ekman 60 Faces Test, based on six facial expressions generated by 10 different people from the Ekman and Friesen (1976) facial expression series (with stimuli kindly made available by Andy Young). The stimuli were selected so that each emotional expression was as well recognised as possible according to the Ekman and Friesen norms. They were presented singly, in a pseudo-random sequence, with the labels 'happiness', 'fear', 'surprise', 'sadness', 'disgust', 'anger' printed underneath each face. PHD was asked 'which word is closest to the emotion being experienced by the person in the picture'. A preliminary series of six expressions was presented (with feedback) followed by 60 test faces, with 10 examples of each emotion, one posed by each person. PHD's total score was worse than that of a sample of 10 control participants (age range 50-68) from the Cambridge subject panel (PHD scored 41/60). He was very impaired on fear, only recognising 2/10 examples (controls: M = 8.6, SD = 1.17). PHD tended to misidentify fear as anger or surprise. He was also significantly impaired on sadness, scoring 6/10 (controls: M = 8.7, SD = 1.34) suggesting disgust, anger and fear as possible options. In the case of anger, he was within one standard deviation of controls, scoring 7 (controls: M = 7.7, SD = 1.42) despite misidentification of sad and fearful faces as angry. He was also within the range of controls for disgust (PHD scored 9; controls: M = 9.0, SD = 1.25); for surprise (PHD scored 8; controls: M = 8.5, SD = 1.58); and was not impaired at recognising happiness (PHD scored 9; controls: M = 9.9, SD = 0.32). Despite typical performance categorising angry faces, PHD was likely to mislabel other emotional faces as angry.

PHD's recognition of emotion was also assessed using body postures (kindly provided by Beatrice de Gelder). PHD was asked to judge whether a posture was happy, sad, angry or fearful. PHD was not as good as the controls reported by de Gelder and Van den Stock (2011), but his score of 20/24 was well above chance.

PHD was also assessed on the TASIT (McDonald, Flanagan, Rollins, & Kinch, 2003): a series of video vignettes that have been designed to evaluate the ability to extract emotional and social cues from short interactions. PHD performed well on those items evaluating positive emotions (happiness, surprise and neutral items) but was at or below the 5th percentile for the negative emotions of anger, anxiety/fear and revulsion. More complex social interactions were mostly understood well. See Eimer and McCarthy (1999) for further details of the abilities of PHD across a variety of measures.

In Experiment 5.1, a control group of four age-matched, males (age range 3-63, M = 52.25, SD = 14.41) were recruited. All participants had normal or corrected-to-normal vision and were asked to complete the Edinburgh Handedness Inventory (Oldfield, 1971) prior to participation. Three were left-handed. No controls had any history of neurological problems.

Stimuli

Grey-scale face morph stimuli were used from the Facial Expression of Emotion test (Young, Perrett, Calder, Sprengelmeyer & Ekman, 2002). Angry, fearful, disgusted and happy faces morphed with neutral were shown (see Figure 5.2). The intensity of the emotion in the morph was varied creating a range of morphs for each emotion with different intensities of emotion relative to neutral. All morphs were based on two images of the same individual. Forty-five morphs were used for each emotion: range 10-98% emotion at 2% intervals (180 morphs in total). Each stimulus appeared at a size of 6.20 cm by 8.80 cm on the screen. Therefore, the visual angle for each image was 7.10° by 10.06°, when viewed on a desktop computer screen from a distance of 50 cm.

Procedure

Participants judged which of two possible emotions was being presented on each trial in six separate blocks (paired combinations of happy, angry, disgusted and fearful). A fixation cross was displayed for 300 ms followed by the stimulus presented centrally on a black background until response. When each stimulus face was removed it was replaced by a rectangle mask of monochrome Gaussian noise displayed for 500 ms. Ninety trials were shown in each condition, with each morph face being shown once. Participants responded by using the mouse buttons in order to select one of the two emotion labels presented on the screen. Condition order and the emotion assigned to each mouse button were counterbalanced for controls and

PHD. Controls each completed the task once and PHD completed the task three times over a period of several months.

See Figure 2 in Mestry, Donnelly, Menneer and McCarthy (2012) due to issues of copyright.

Figure 5.2. Examples of morphed face stimuli shown in Experiment 5.1. Angry, disgust, fear and happiness (left to right). Percentage of emotion morphed with neutral: 10%, 50% and 90% (top to bottom).

Results

The data were aggregated into four intensity ranges: 10-30%; 32-52%; 54-74%; and 76-98%. These aggregated data were converted into sensitivity (d') and bias (c) scores. These data for controls were analysed in two separate 6 (emotion-pair) x 4 (intensity) repeated measures ANOVAs. Separate ANOVAs were performed for d' and c.

For sensitivity (d'), there was a main effect of emotion-pairing ($F(5,15) = 4.56$, $p = 0.010$). Sensitivity was highest for the fear and disgust comparison ($M = 3.13$, $SE = 0.22$) and lowest for the anger and fear comparison ($M = 2.11$, $SE = 0.36$) but pairwise comparisons revealed there were no significant differences in sensitivity between the six emotion pairings. There was also a main effect of intensity ($F(3,9) = 27.88$, $p < 0.001$). Sensitivity improved linearly from the 10-30% to the 54-74% intensity level (see Figure 5.3) and reached asymptote at this point with pairwise comparisons revealing significant differences between the 10-30% intensity range and the three other intensity ranges (32-52%, $p = 0.026$; 54-74%, $p = 0.027$; and 76-98%, $p = 0.033$), but not amongst any other comparisons (ps > 0.517). The interaction of category and intensity for sensitivity was not significant ($F(15,45) = 1.10$, $p = 0.39$). There were no significant effects of bias (c) (Fs < 0.02, ps > 0.996).

PHD was compared to controls using Crawford and Howell's (1998) method which uses the control sample statistics as statistics in the test rather than as parameters, which is more appropriate when the control sample is modest in size. The Siglims_ES.exe program, made available by Crawford, Garthwaite and Porter (2010), was used along with Bonferroni correction to correct for multiple comparisons. Significant differences at $p = 0.01$ are indicated in Figure 5.3. These results demonstrated that PHD performed as controls except when discriminating fearful from disgusted faces in the 54-74% and 76-98% ranges ($t(3) = -184.88$, $p < 0.001$; $t(3) = -111.50$, $p < 0.001$), and angry from disgusted faces ($t(3) = -195.43$, $p < 0.001$) and happy from angry faces ($t(3) = -80.77$, $p < 0.001$) in the 54-74% range where his d' was significantly lower than for controls. Also discriminating happy from angry ($t(3) = 29.16$, $p < 0.001$) and happy from disgust faces ($t(3) = 29.16$, $p < 0.001$) in the 76-98% range showed PHD's d' was significantly higher than for controls. Note the values for c in Figure 5.3 are very small (indicating no bias).

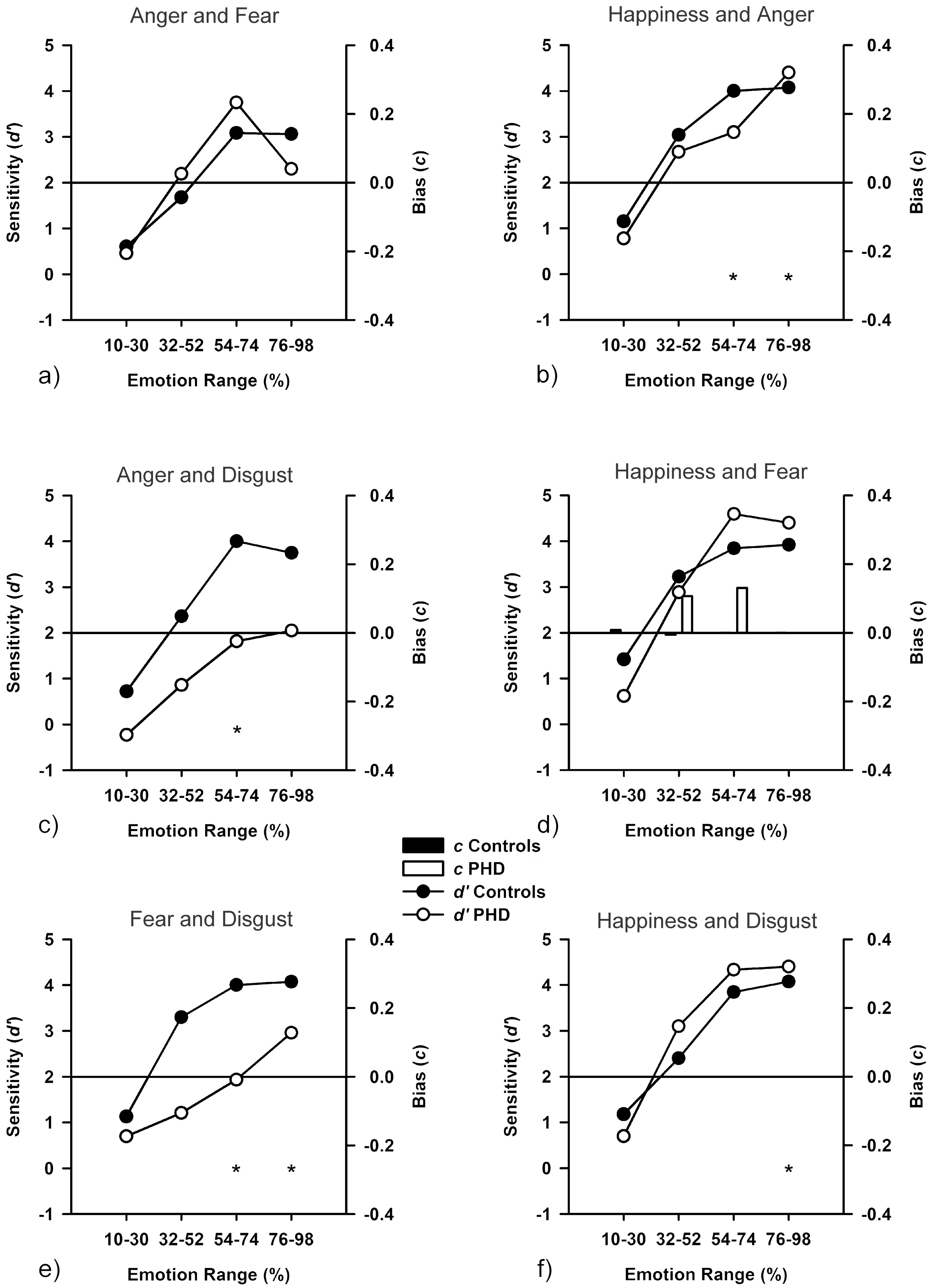


*Figure 5.3. Sensitivity (d') and bias (c) data for PHD and controls in all pair-wise categorisations in Experiment 5.1 with * to indicate significance at p = 0.01 (two-tailed): (a) anger and fear, (b) anger and happiness, (c) anger and disgust, (d) fear and happiness, (e) fear and disgust and (f) happiness and disgust. Note, bias values are very small and therefore appear absent in most figures (indicating no bias).*

Discussion

The clinical data showed PHD to be poor at categorising fear and sadness, and in some respects angry if misidentification is considered. PHD's impoverished categorisation of these particular emotions may reflect an absence of configural processing. However, these clinical data are difficult to interpret. First, it is possible that performance reflects a difficulty using such information in the context of a six-alternative forced choice decision. Second, the categorisations may be subject to criterion shifts, such as overestimating the relative frequency of some emotions, e.g. angry faces.

To overcome these issues PHD was tested in conditions where the alternatives were limited to two and sensitivity could be determined independently of bias. These data show that, given sufficient intensity, discriminating between pairs of emotions is rather easy for control participants. More importantly, the data show that PHD can make pair-wise categorisations of facial emotions, especially if one of the face categories is happiness. However, PHD is markedly less sensitive than controls when categorising anger from disgust and fear from disgust, especially at high levels of stimulus intensity.

By themselves, these data can be interpreted in a number of different ways. The data were explored further in a follow-up study: PHD and controls performed a psychophysical intensity discrimination threshold task using the same face set as in Experiment 5.1. I sought to establish the magnitude of the difference in intensity between two faces of the same emotion that was required before PHD and controls could reliably report faces of the higher intensity. The reasoning is that uncertainty in relation to faces of specific emotions would translate into high discrimination thresholds (a higher percentage difference in emotion required to make the discrimination). Using a one-up, two-down threshold paradigm, the point at which face intensity could be reliably discriminated on 71% of occasions was measured. In this staircase paradigm, correct discrimination leads to a reduced difference in emotion intensity between the stimuli on the next trial, whilst incorrect discrimination leads to increased difference. The threshold for reliably discriminating between pairs of simultaneously presented angry, happy, fearful and disgusted faces was measured. Again, Crawford and Howell's (1998) method for comparing a single participant to a small control group with Bonferroni correction was used. The data show PHD to be significantly poorer than controls at within-emotion intensity judgements of anger (t(3) = 9.22, p = 0.001) and disgust (t(3) = 8.934, p = 0.001), but not of happiness (t(3) = 2.32, p = 0.103) or fear (t(3) = 1.77, p = 0.174, see Figure 5.4).

Together, the data from Experiment 5.1 suggest PHD to have difficulties in the perception of anger and disgust, some difficulty in the categorisation of anger from fear and no difficulty in the perception or categorisation of happiness. This pattern of results is consistent with a loss of configural information, where some emotions will be more affected than others (McKelvie, 1995; Calder et al., 2000).

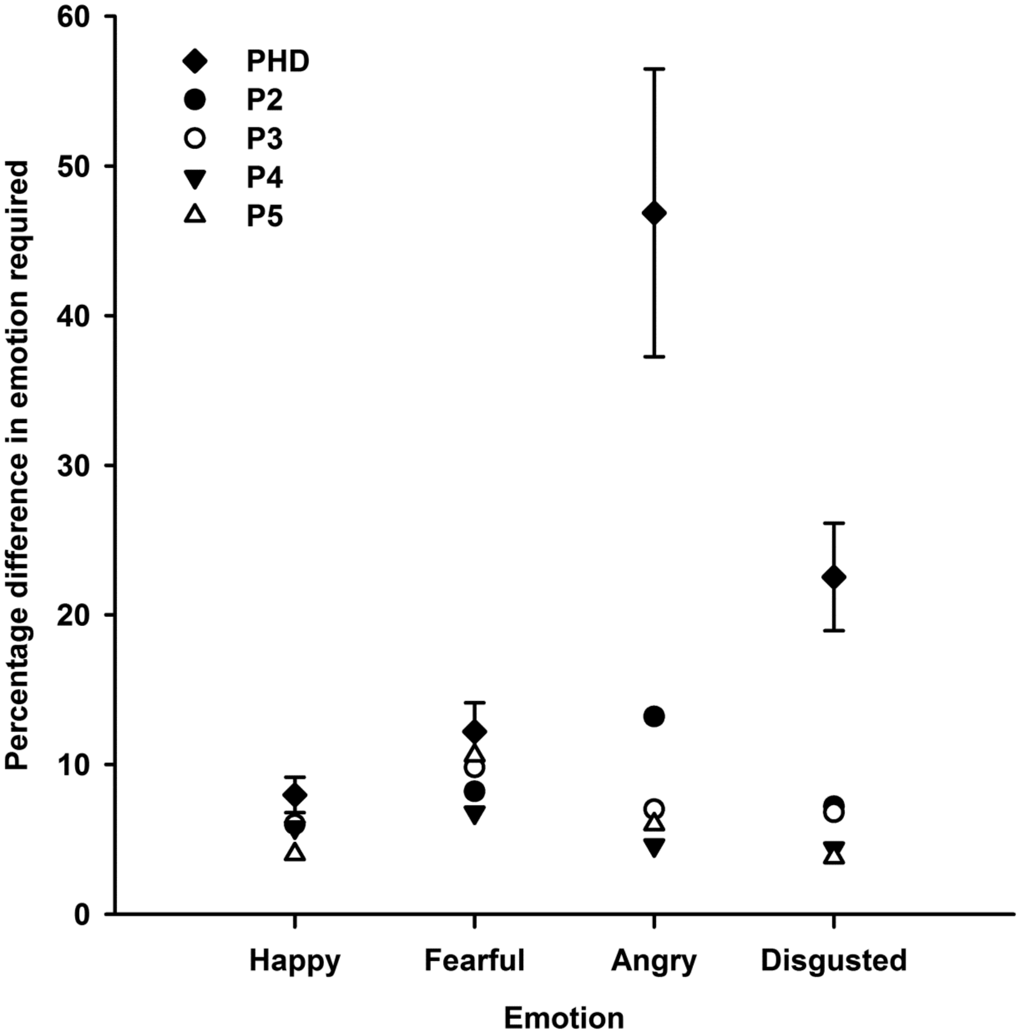


Figure 5.4. Percentage difference in emotion content required to discriminate magnitude of emotion for PHD and each control participant (P2-P5). Values for PHD represent the mean difference across four repeats and therefore, include standard error bars of the mean.

Experiment 5.2: Perception of the Thatcher Illusion

In Experiment 5.2a, categorisation of Thatcherised from typical faces in PHD and controls was explored. Using a task similar to that used in Experiment 5.1, participants were asked to determine if individually presented stimuli were Thatcherised faces or typical faces. Given the salience of the Thatcher illusion for typical participants, it was anticipated that controls would perform as if discriminating between highly salient emotions. However, this assumption needs to be stated more formally. If it is assumed that the d's of Experiment 5.1 reflect the perceived difference in emotional valence of face categories, then the d's of controls when categorising Thatcherised from typical faces can be matched with those from Experiment 5.1 when categorising emotional faces. The critical question is how PHD performs in the condition that has been matched for valence. In particular, can his performance in Experiment 5.2a be predicted from his performance in Experiment 5.1?

All participants were run in an additional study (Experiment 5.2b). In Experiment 5.2b, participants made 'same' and 'different' decisions to pairs of faces composed of orthogonal combinations of Thatcherised and typical faces shown in Experiment 5.2a. In addition to testing participants in Experiments 5.2a and 5.2b with faces, participants were also tested using images of churches that were manipulated (Thatcherised) in the same way as faces.

Method

Participants

PHD and eight left handed male controls were tested. Two controls were age-matched (aged 54 and 56) and six were students (age range 20-29, M = 23.17, SD = 3.66). All participants had normal or corrected-to-normal vision and were asked to complete the Edinburgh Handedness Inventory (Oldfield, 1971) prior to participation; all were left-handed. No controls had any history of neurological problems.

Stimuli

Ten grey-scale face stimuli from the NimStim face set (Tottenham et al., 2009) were used to create Thatcherised stimuli by inverting the eyes and mouths. Grayscale images of churches were used as control stimuli and were manipulated in a similar way to faces by inverting the windows and the door (see Figure 5.5). Individual face stimuli appeared at a size of 10 cm by 13 cm and individual church stimuli at a width of 8.30cm, although the height varied between 7.59 cm and 16.98 cm. Therefore, the visual angle was 7.63° by 9.91° for faces, and 6.33 by between 5.79° to 12.92° for churches, when viewed on a desktop computer screen from a distance of 75cm.

Figure 5.5. Examples for face and church stimuli shown in Experiment 5.2a (Categorisation Task) and Experiment 5.2b (Discrimination Task).

Categorisation Task (Experiment 5.2a)

Participants were instructed to decide if the stimulus (either a church or a face) was ‘typical’ or ‘odd’. Odd was defined by explaining how they had been changed to look grotesque and by showing examples of Thatcherised faces versus typical faces (and the equivalent versions for churches). After a 250 ms fixation cross, stimuli were presented centrally until a response was made. The experimental design incorporated 10 individual faces, each in a Thatcherised and typical form. Upright and inverted versions of each were repeated four times, creating 160 trials. The same design was used for churches, creating 160 church trials. The order of stimulus type was counterbalanced across participants. Controls completed the task once and PHD completed the task three times over a period of several months.

Discrimination Task (Experiment 5.2b)

Participants were instructed to decide whether pairs of stimuli were 'the same' or 'different'. After a 250 ms fixation cross, stimuli were displayed until a response was made by pressing one of two designated mouse buttons. The stimuli from the Categorisation Task were combined into pairs and shown centrally, separated by a 2 cm gap for simultaneous comparison. Equal numbers of matching and mismatching pairs were created using the same face/church identity and orientation, with the only difference in a mismatched pair being that one stimulus was Thatcherised and one was typical. Each stimulus appeared eight times in equal numbers of upright and inverted presentations. There were a total of 160 face comparison trials and 160 church comparison trials. Trials were randomised and blocked by object type, and block order was counterbalanced across participants. Controls completed the task once and PHD completed the task three times over a period of several months.

Results

As in Experiment 5.1, all data were analysed in terms of signal detection measures, d' and c. The age-matched controls were always within the range of student controls and therefore the data from all control participants was combined. The results were analysed using 2 (stimulus: faces versus churches) x 2 (orientation: upright versus inverted) repeated measures ANOVAs. Separate analyses were computed for the Categorisation and Discrimination Tasks as well as for d' and c.

The main effect of orientation was significant for d' in both the Categorisation and Discrimination tasks ($F(1,7) = 57.64$, $p < 0.001$; $F(1,7) = 11.02$, $p = 0.013$ respectively). Participants were more sensitive to upright ($M = 3.03$, $SE = 0.14$; $M = 4.24$, $SE = 0.26$) than inverted stimuli ($M = 2.34$, $SE = 0.13$; $M = 3.80$, $SE = 0.26$). The main effect of stimulus type was significant in the Categorisation Task ($F(1,7) = 6.34$, $p = 0.040$) with sensitivity higher to churches ($M = 2.94$, $SE = 0.17$) than to faces ($M = 2.43$, $SE = 0.16$). There was no significant effect of stimulus type in the Discrimination Task ($F(1,7) = 3.20$, $p = 0.117$).

Orientation and stimulus type yielded a significant interaction in the Categorisation Task thus replicating many previous studies of the Thatcher illusion ($F(1,7) = 13.55$, $p = 0.008$). Inversion reduced sensitivity to faces in the categorisation task (upright, M = 3.15, SE = 0.16; inverted, M = 1.72, SE = 0.25), but there was no difference for churches (upright, M = 2.92, SE = 0.15; inverted, M = 2.95, SE = 0.20). There was no significant interaction between orientation and stimulus type in the Discrimination Task ($F(1,7) = 4.41$ $p = 0.074$).

With respect to the c data, there was a main effect of stimulus type on the Categorisation Task ($F(1,7) = 7.08$, $p = 0.032$), with bias towards the 'odd' response being greater for faces (M = -0.27, SE = 0.09) than for churches (M = -0.02, SE =0.09). There was no main effect of stimulus type on bias in the Discrimination Task ($F(1,7) = 0.00$, $p = 0.999$). The main effect of orientation was not significant in either the Categorisation or the Discrimination Tasks ($F(1,7) = 4.09$, $p = 0.08$; $F(1,7) = 0.82$, $p = 0.396$). However, there was an interaction between stimulus type and orientation in both the Categorisation and Discrimination Tasks ($F(1,7) = 16.30$, $p = 0.005$; $F(1,7) = 10.03$, $p = 0.016$ respectively). Responses to faces were more biased towards 'odd' and 'different' when inverted (M = -0.60, SE = 0.16; M = -0.20, SE = 0.15) than when upright (M = 0.07, SE = 0.08; M = 0.07, SE = 0.08). In contrast, the reverse was true with churches for upright (M = -0.08, SE = 0.09; M = -0.12, SE = 0.07) and inverted (M = 0.04, SE =0.14; M = -0.01, SE = 0.12) respectively.

PHD was compared against chance ($d' = 0$) and the pooled control groups using one sample t-tests with Bonferroni correction. PHD was above chance in all conditions of the Discrimination Task ($ts(2) > 9.76$, $ps < 0.010$). In the Categorisation Task he was above chance in the upright ($t(2) = 15.05$, $p = 0.004$) and inverted ($t(2) = 5.96$, $p = 0.027$) church conditions but he was unable to perform the face task scoring at chance for both the upright ($t(2) = 0.20$, $p = 0.858$) and inverted ($t(2) = -0.37$, $p = 0.746$) face conditions (see Figure 5.6). When comparing PHD's sensitivity to controls (Crawford & Howell, 1998, with Bonferroni correction), there were no significant differences in any condition of the Discrimination Task (magnitude of $t(7)s < 1.42$, $ps > 0.198$) or the upright and inverted church conditions of the Categorisation Task ($t(7) = -1.03$, $p = 0.337$; $t(7) = -1.97$, $p = 0.090$). However, PHD's inability to categorise faces was evident, with controls significantly more sensitive than PHD in the upright face condition of the Categorisation Task ($t(7) =-6.38$, $p < 0.001$), and marginally more significant in the inverted condition ($t(7) = -2.28$, $p = 0.056$).

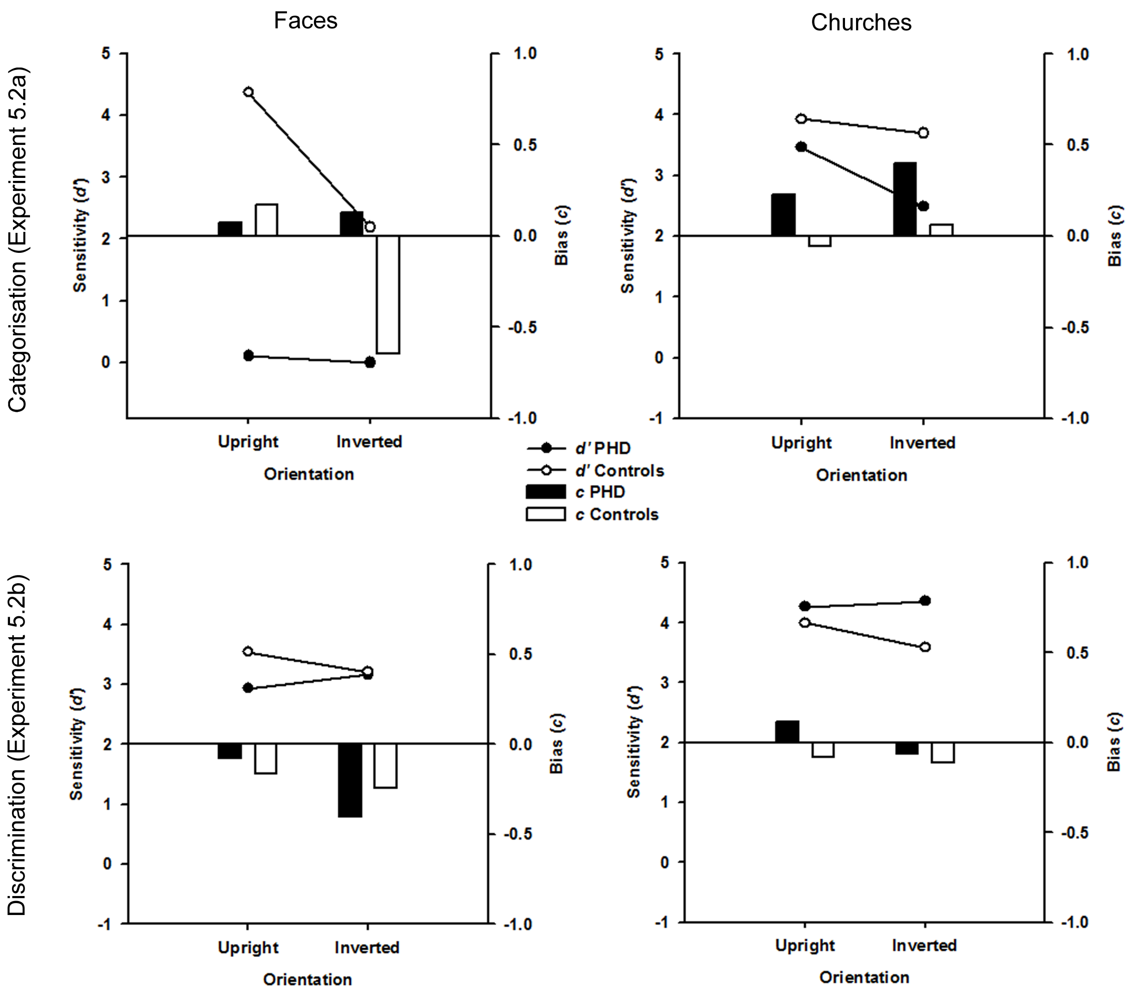


Figure 5.6. Sensitivity (d') and bias (c) data for face (left) and church (right) conditions in Experiment 5.2a (top) and Experiment 5.2b (bottom).

Discussion

The Categorisation Task of Experiment 5.2a revealed that controls showed, as expected, good sensitivity to upright and inverted faces and churches. The sensitivity of controls to upright faces was similar to the levels achieved when categorising highly intense emotions in Experiment 5.1. Although equally sensitive when categorising upright faces and churches, controls were more sensitive when categorising inverted churches compared to inverted faces. In contrast, PHD performed at chance-level categorising both upright and inverted faces. However, this was not evident for churches, where his performance was above chance and no different to controls. It might have been that PHD was unable to perceive the stimulus alterations that differentiated Thatcherised from typical faces. However, the simultaneous discrimination task of Experiment 5.2b shows this to not be the case. PHD was sensitive to differences in both faces and churches when both upright and inverted.

Comparing across Discrimination and Categorisation Tasks, the key contrast is that PHD cannot categorise Thatcherised from typical faces, despite being able to discriminate the very same faces when they are shown simultaneously.

General Discussion

Controls were sensitive to categorising facial emotions, discriminating intensity of emotional expressions, and categorising and discriminating Thatcherised from typical faces. They also demonstrated the expected inversion effect for faces when categorising Thatcherised from typical faces. PHD's results differ from controls' in several important ways. First, while PHD can perform pair-wise categorisations of emotional expressions, providing emotions are sufficiently salient, he is poor at judging relative intensity within some categories (anger and disgust). Second, PHD is at chance categorising Thatcherised from typical faces, despite being able to discriminate differences between these faces and being able to categorise churches.

Previously, I suggested that this conjunction of findings is consistent with an absence of configural face processing. However, there are two competing hypotheses. Firstly, that Thatcherised faces should be thought of as variants of angry faces (as suggested by a reviewer of Mestry, Donnelly et al., 2012). This was suggested due to PHD's impaired with angry faces in Experiment 5.1, and failure with Thatcherised faces in Experiment 5.2a. Secondly, the Thatcherised faces shown in Experiment 5.2 were less intense in their perceived emotion than the emotional faces shown in Experiment 5.1. If so, it might be that PHD's failure in Experiment 5.2 reflects a mere intensity effect.

With respect to the hypothesis that Thatcherised faces are closely matched to angry faces, PHD was asked to categorise angry, neutral and Thatcherised faces with a view to exploring the resulting confusion matrix. PHD was not very sucessful at this task (78.13% correct on neutral trials, 58.13% on angry face trials and 62.50% correct on Thatcherised trials). More importantly, he mistook Thatcherised faces as neutral faces (24.38%) more often than as angry faces (13.13%). These results suggest that PHD does not see Thatcherised faces as angry faces and that his difficulties perceiving both classes of face, although related, are actually separate issues.

With respect to the hypothesis that the Thatcherised faces shown in Experiment 5.2 were less intense than the emotional faces shown in Experiment 5.1, the salience of faces across Experiments 5.1 and 5.2 were compared directly. For controls, the *d*'s in Experiment 5.2 matched that of high and high-medium emotion conditions in Experiment 5.1. Therefore, PHD's failure to categorise Thatcherised faces from typical faces cannot result from Thatcherised faces being of low intensity.

In conclusion, these findings are consistent with PHD being able to use facial features to map some emotional faces into certain emotion categories. However, doing this requires being able to compute the similarity of actual facial features to those that define each category. At the limit, featural similarity can be estimated in terms of the similarity of pixels across images in a way that is not true for similarity determined by configural relations. If PHD categorises faces (Thatcherised and emotional) with reference to simple features then featural similarity between the sets of images forming the categories should predict his behaviour in a way not true for controls.

This idea was tested by computing the similarity between pairs of sets of images used in the categorisation tasks of Experiments 5.1 and 5.2. The sets of images were determined by the sets over which sensitivity was calculated. For example, sensitivity from the emotion task was calculated between the set of low intensity (10% to 30%) happy faces and the set of low intensity angry faces. There were 16 sets altogether (four emotion ranges for four different emotions). For the Thatcher task, sensitivity was computed between the set of typical faces and the set of Thatcherised faces, providing two sets of images (typical and Thatcherised). For each set of images, an average image was created by taking the mean of RGB values for each pixel location across all images in the set of faces. This average image formed a single representation of all images in the corresponding set. The difference between a pair of average images was then computed by taking the difference in RGB values at each pixel location, summing the squares of these differences, and taking the square root of the sum to provide a measure of dissimilarity between the pair. Therefore, these dissimilarity measures therefore represented the overall difference between two sets of images. These dissimilarity data were then regressed against the mean sensitivity data for all pair-wise categorisation decisions.

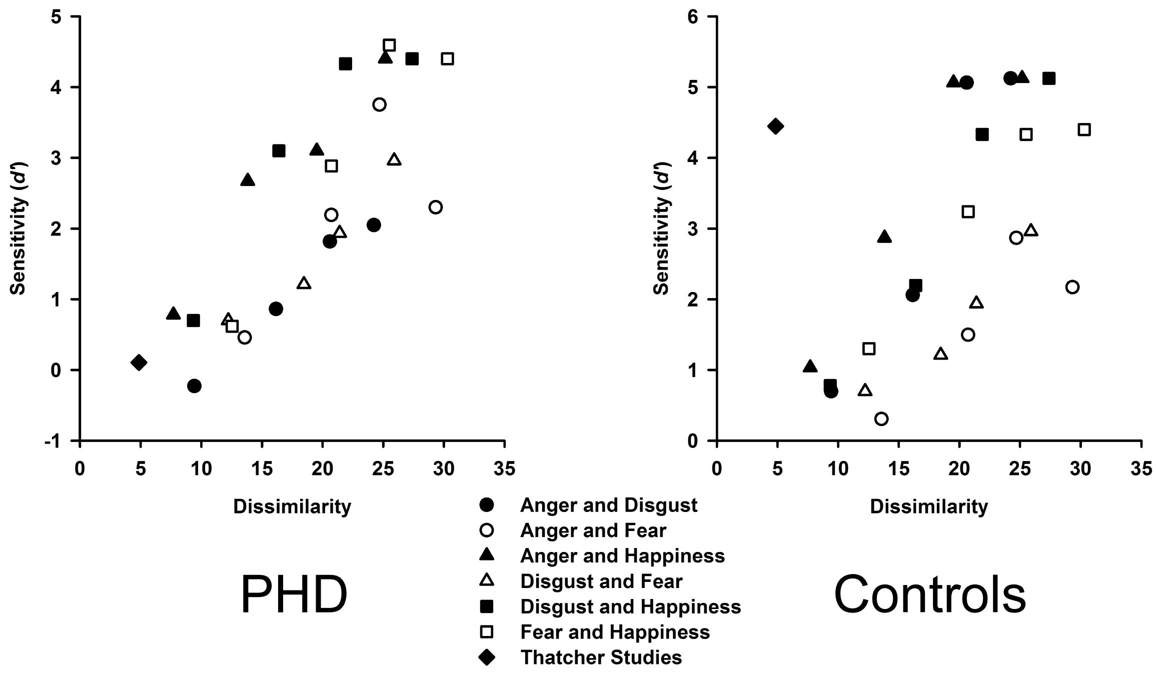


Figure 5.7. Graphs showing the regression of d' against image similarity for PHD (left) and controls (right), using the d' values from Experiment 5.1 and the upright face conditions in experiment 5.2a.

The results show that PHD's sensitivity, across all categorisation conditions of Experiments 5.1 and 5.2, is predicted by this simple feature similarity model (see Figure 5.7). PHD was more sensitive to categories when the images drawn from each category were physically very different. His sensitivity reduces as the difference between images reduces. The regression model was significant, $F(1,23) = 42.22$, $p < 0.001$. Similarity is a significant predictor of sensitivity ($\beta_1 = 0.17$, $t(23) = 6.50$, $p < 0.001$, $R^2 = 0.65$). In contrast, whilst the regression model is also significant for controls ($F(1,23) = 12.77$, $p = 0.002$ with similarity as a significant predictor of sensitivity, $\beta_1 = 0.15$, $t(23) = 3.57$, $p = 0.002$, $R^2 = 0.38$), the data emphatically show that control participant's responses to Thatcherised faces are not based on simple feature similarity, as is the case for PHD (see Figure 5.7). Controls have an exquisite sensitivity to the manipulations used to create Thatcherised faces.

These data suggest that, for controls, the categorisation and discrimination of Thatcherised from typical faces results from the computation of some configural feature around eyes and mouths, or from the difficulty of attempting to compute such features. These two classes of explanation were raised in the previously, but there is one further point to be added: the current data cannot determine which of these accounts is correct. Other papers have explored the nature of configurality in the Thatcher illusion, suggesting an absence of configural processing in upright Thatcherised faces (Donnelly, Cornes, & Menneer, 2012; Mestry, Menneer, Wenger & Donnelly, 2012 (Experiment 2.1)). Therefore, these data would suggest that it is PHD's absence of configural processing in typical faces, alongside the general absence of configural processing of Thatcherised faces, which leaves him unable to categorise Thatcherised from typical faces.

Conclusion

This study has shown that discrimination of Thatcherised and emotional faces is possible in a patient who lacks the N170. However, the ability to categorise Thatcherised faces is lost. A deficit in configural processing impacts PHD's categorisation of Thatcherised faces and to some extent his emotion processing while his preserved feature processing supports his ability to categorise emotional faces and his ability to discriminate between Thatcherised and typical faces. These results support the role of configural processing in both perception of the Thatcher illusion and typical emotion processing, but cannot speak to sources of configurality being present in the upright Thatcherised face. The Thatcher illusion is considered to be a face specific illusion where configural processing is dependent on the automatic categorisation of faces (see Boutsen, Humphreys, Pramstra, & Warbrick, 2006; Milivojevic, Clapp, Johnson, & Corballis, 2003; Carbon, Schweinberger, Kaufman, Leder, 2005 for explorations of the N170 in relation to Thatcher illusion). PHD does not show an N170 face effect (Eimer & McCarthy, 1999) and does not experience the oddity of Thatcherised faces. Also, the results of Experiment 3.1 reveal the effect of Thatcherisation is at the N170 component. Therefore, these results suggest that the neural substrate for automatic face categorisation, indexed by the N170, is a prerequisite for the experience of discomfort and oddity that characterises neurotypical individual's encounters with Thatcherised faces.

Prosopagnosia and the Thatcher Illusion: A General Recognition Theory Account

My previous study using GRT (Experiment 2.1, Chapter 2) suggested that there is evidence for configurality in the Thatcher illusion from multiple sources, and that configural processing is evident for inverted faces too. Also, Cornes, Donnelly, Godwin and Wenger (2011) showed that the effect is driven by decisional factors (violation of decisional separability, DS) and not just by perception as previously assumed. Finding violations of perceptual independence (PI), perceptual separability (PS) and DS in different combinations across participants for a Thatcher illusion task suggested that there is more than one type of configurality involved in perception of the Thatcher illusion. To understand the nature of the illusion, a GRT task with a prosopagnosic individual was explored to assess whether there would be any residual evidence for one or more types of configurality in the same task. There is more than one type of prosopagnosia and each leads to a different pattern of deficits. It could be hypothesised that different types of prosopagnosia are related to deficits in different types of configurality, identified by the violations of GRT concepts.

Experiment 5.3 Pilot GRT Studies

PHD is an acquired apperceptive prosopagnosic, and previous studies have shown he has other associated deficits that make traditional task designs quite challenging. Therefore, there may be an issue with the suitability of the GRT framework for investigating perception of the Thatcher illusion with this individual. Experiments 5.1, 5.2 have shown PHD relies on feature based processing to perform in face processing tasks. The challenge was to try to identify a complete factorial methodology in which PHD may show some residual configural processing, but controls could also perform to provide comparative data to show the task does require the use of configural processing.

Aim

(1) To assess the suitability of the GRT paradigm to explore face processing abilities with PHD.

The first task PHD attempted was the face condition of the Cornes et al. (2011) factorial Thatcher task. PHD found the stimuli appeared too fast and he was unable to successfully perform at the task where multiple responses were required; thus this experiment was not pursued further (Personal Commentary, February, 2008). PHD was then tested on the same task used with children and adults in Chapter 4 (Experiment 5.4a). In this task, face stimuli were shown for 1000 ms, much longer in comparison to the 100 ms exposure duration used in the Cornes et al. (2001) paradigm.

Experiment 5.3a: Developmental Thatcher Illusion Task

PHD completed one block of upright trials and one block of inverted trials (40 of each) in the developmental Thatcher task (see Experiment 4.1, Chapter 4), the equivalent completed by each participant in this task. He again reported finding the 1000 ms exposure duration too fast and that he was guessing to make responses and so no further blocks were run. Despite there being insufficient data to analyse using GRT, some comment on his performance is possible. PHD performed at a level lower than for the youngest age group of 8/9 year olds (see Figure 5.8). However, he does show residual reduced sensitivity to the inverted compared to upright condition, and shows increased bias to report the eyes as 'normal' in the upright version of the task.

PHD's performance was compared formally using one sample t-tests with Bonferroni correction (Crawford & Howell, 1998). The d' and c results for PHD in each of the upright and inverted conditions were compared to each of the three age groups. PHD did not perform any differently to the controls except for sensitivity to upright faces where adult controls (M = 2.15) were more sensitive than PHD (M = 0.47, $t(9) = -2.92$, $p = 0.017$).

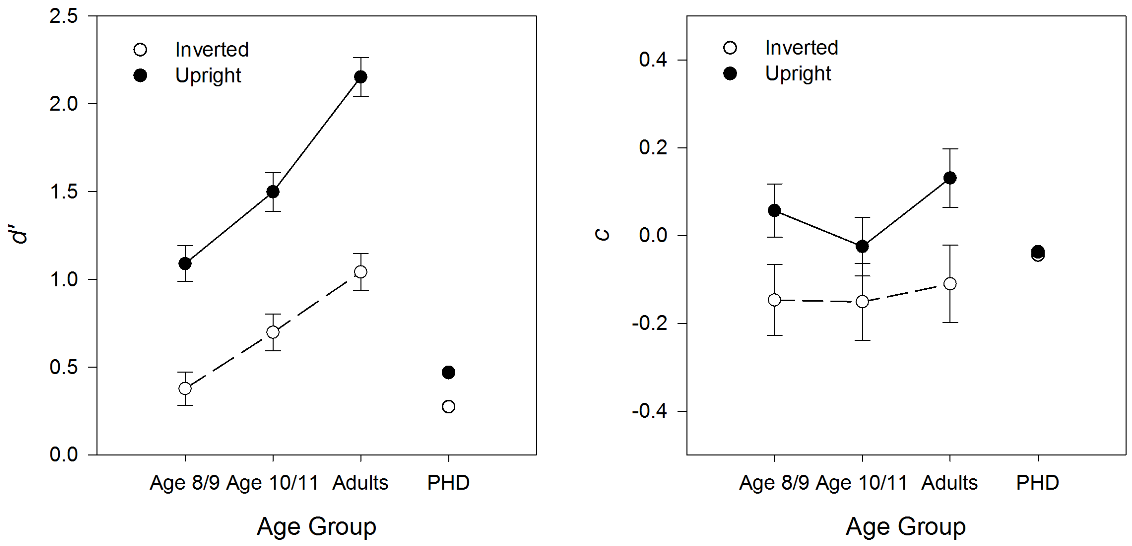
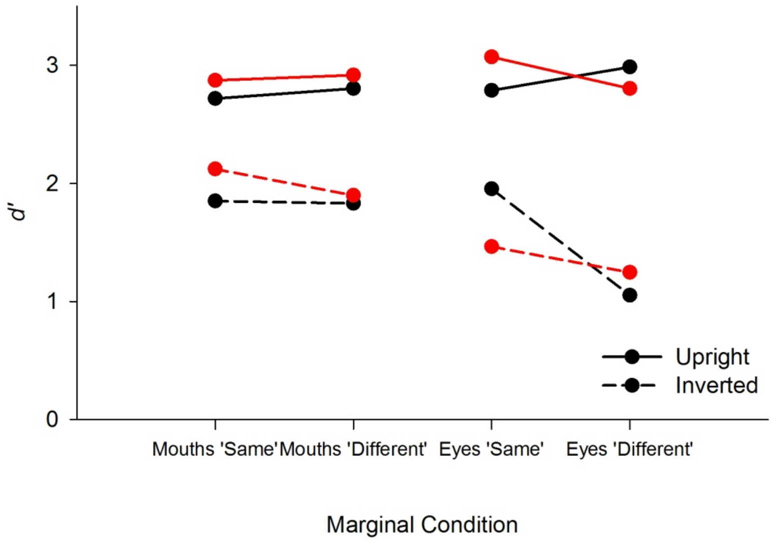


Figure 5.8. Graphs showing d' (left) and c (right) in the upright and inverted versions of the developmental Thatcher task for typical participants in Experiment 4.1 compared to PHD.

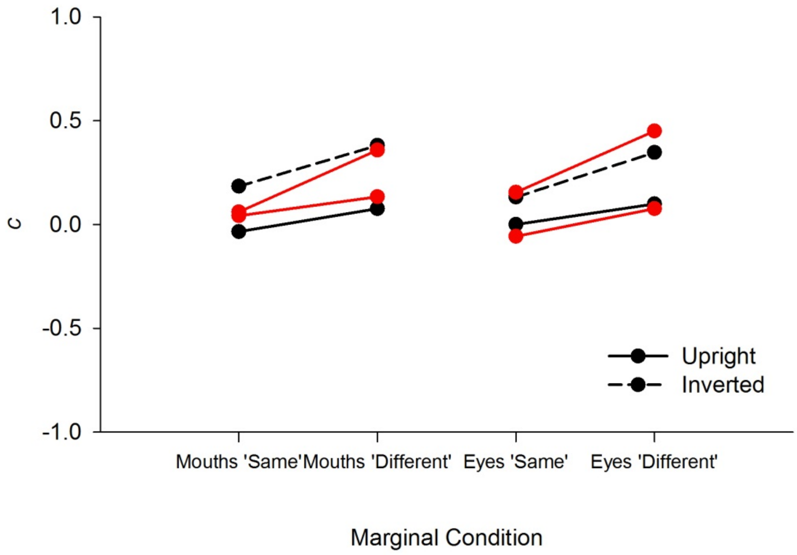
Experiment 5.3b: Same-Different Feature Matching Task

Having shown PHD to be unable to perform in a full factorial experiment using brief presentations of Thatcherised faces, the task was simplified for an unlimited exposure duration and simultaneous matching. The stimuli used were those outlined in Experiment 3.1, and used in Experiments 4.1 and 5.3a. Pairs of faces of the same identity were presented and the task required participants to judge the status of two sets of features (the eyes and mouths) across two levels (same or different). Together, these two dimensions, each with two levels, created four stimulus conditions. Every possible pairing of the four face conditions for each identity (i.e. typical, eyes Thatcherised, mouth Thatcherised, both eyes and mouth Thatcherised) were shown to create the same number of same and different trials: DD, DS, SD and SS. There were two versions of the task, one with upright faces and the other with inverted faces, these were completed as separate blocks but trials within blocks were randomised. A 500 ms fixation cross was followed by a pair of stimuli displayed on the screen. A response was required for each feature and the pair of stimuli was displayed until both responses were made. The orientation version of the task, the order of feature to respond to, and the button responses were counterbalanced within and between participants.

PHD completed 120 trials, 30 for each stimulus type (see black lines, Figure 5.9). Two age matched controls completing the same task (60 year old males who had previously taken part in Experiments 5.1 and 5.2) with a total of 160 trials between them, 40 for each condition. All participants were at ceiling as the task is trivial and can be completed using a feature based strategy and it does not require any face representation to complete the task. Some element of configurality may contribute to performance but if the task promotes a featural strategy then it cannot be used to compare configural face processing of PHD to typical individuals.



A



B

Figure 5.9. Graphs showing (A) marginal sensitivity (d') and (B) marginal bias (c) values for PHD and age matched controls in the Same-Different Feature Matching Task. Black lines show values for PHD, red lines show values for the age matched controls. The marginal conditions ar:e eyes when mouths are the same, eyes when mouths are different, mouths when eyes are the same and mouths when eyes are different.

Experiment 5.3c: Geometric Shape Task

Having shown PHD is unable to perform in a full factorial experiment using brief presentations of Thatcherised faces, the task methodology was simplified whilst maintaining suitability for the GRT framework. Faces were made using geometric shapes (Figure 5.10a). The ‘eyes’ could either be circles or ovals and the ‘mouth’ could either be a square or diamond. Stimuli had set exposure durations and presentation order was randomised. A 500 ms fixation cross was followed by a 100 ms mask (Figure 5.10b), a stimulus for 500 ms, then the mask again for 100 ms before the two responses were required. There was no variation in stimuli within each condition.

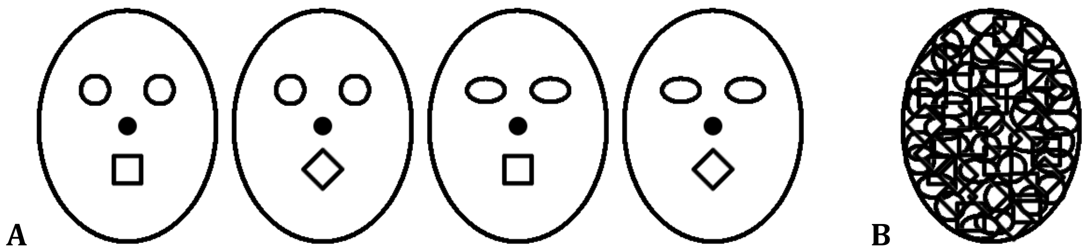


Figure 5.10. (A) stimuli and (B) mask used in the Geometric Shape Task.

PHD completed 400 trials, 100 for each stimulus type and was at chance identifying each feature (see black filled shapes, Figure 5.11). Compared to an age matched control (a right handed, 60 year old male who had previously taken part in Experiments 5.1 and 5.2 , and 5.3b) who completed the same task with an 80 ms stimulus exposure, it is clear PHD is unable to perform at this task even when core task was simplified. PHD has a sensitivity close to zero for all marginal conditions. There is no variation in features (always the four ‘faces’) and PHD has knowledge of the prototype shapes.

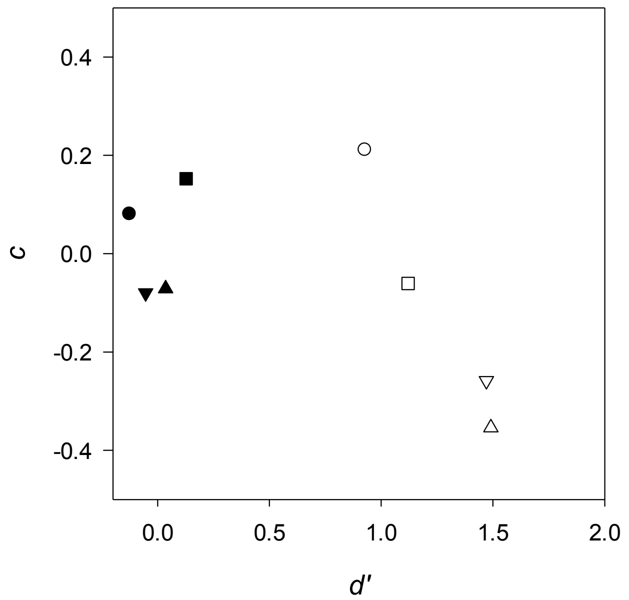


Figure 5.11. Graph showing the marginal sensitivity (d') and bias (c) values for PHD and an age matched control in the Geometric Shape Task. Black shapes show values for PHD, white shapes show values for the age matched control. The marginal conditions are: circle = top when bottom is square, square = top when bottom is diamond, up pointing triangle = bottom when top are circles and down pointing triangle = bottom when top are ovals.

Conclusion

An apperceptive prosopagnosic, referred to as PHD, was unable to perform in the Thatcher illusion task within the GRT framework used for typical participants in Chapter 4. There was some evidence for a residual inversion effect using the task (Experiment 5.3a). However, in a simplified version with no feature variation and limited exposure duration (Experiment 5.3c), PHD was still unable to perform the task. Therefore, the evidence from Experiment 5.3a is unlikely to represent residual configural processing and is more likely to be an artefact from a small sample of trials. PHD can perform discrimination tasks with unlimited exposure duration (Experiment 5.3b), but these tasks do not require any representation of configural information and can be completed featurally. Together, these results suggest PHD is unable to hold a mental representation of the stimuli, whether they are faces or objects. Overall, it was found that a GRT paradigm could not compare the face processing abilities of PHD to controls. For PHD, the GRT paradigm is unsuitable, but for other prosopagnosics it could be valuable to determine the precise nature of the deficits and to discover whether there is any evidence for residual configural processing.

Chapter 6

General Discussion

The motivation of this thesis was to explore configural face processing. In Chapter 1, several experimental phenomena that are used in support of theories of configural face processing were reviewed. I have demonstrated (1) inversion does not exert a qualitative influence on face processing, (2) differences in holistic processing between faces and objects are quantitative rather than qualitative, and (3) effects attributed to perceptual representation are sometimes driven by other sources of influence (e.g. emotion processing). Most importantly, I described recent studies that have highlighted high-level influences of bias in the part-whole effect, the composite face effect and the Thatcher illusion (e.g. Wenger & Ingvalson, 2002, 2003; Richler, Gauthier, Wenger, & Palmeri, 2008; Cornes, Donnelly, Godwin, & Wenger, 2011). Together these findings indicate the need to re-evaluate previous theories. The primary inference of this work is that prior interpretation of studies reflects an incomplete mapping between data and theory. To address the situation, I have reworked three classic face processing tasks along the lines of three previous studies (Wenger & Ingvalson, 2002, 2003; Richler et al., 2008; Cornes et al., 2011) such that the data could be analysed using general recognition theory (GRT). Studies in later chapters explored one of these tasks in detail.

Findings, Implications and Limitations

Chapter 2

In Chapter 2, three face processing tasks were examined for evidence of configurality. Configurality was defined mathematically as three types of violations outlined by GRT (e.g. Ashby & Townsend, 1986; Kadlec & Hicks, 1998; Kadlec & Townsend, 1992a, 1992b; Mack, Richler, Gauthier, & Palmeri, 2011). Violations represent perceptual and decisional dependences between facial features; the relationship between these violations and configurality has been investigated in previous papers (see Wenger & Ingvalson, 2002).

Before summarising the data in terms of the violations defined by GRT, it is important to note that the data from Chapter 2 replicate the results of many previous studies when considered as condition means analysed using standard statistics (ANOVA). Specifically, inversion effects were found for all three tasks. Inversion effects were present in the sensitivity data for the Feature Identity and Feature Orientation Tasks and in the bias values for all three tasks. Higher sensitivity was found with upright compared to inverted faces. Also, participants found it harder to report difference in the Feature Size and Feature Identity Tasks, as well oddity for the Feature Orientation Task when faces were inverted as opposed to upright. Within the context of a standard experiment measuring RT or accuracy, these effects are likely to lead to lower accuracy and slower RTs when shown inverted rather than upright faces. For composite face studies, there is a difficulty demonstrating behavioural effects in the same way because two responses are required instead of one, and a complete as opposed to a partial design is used (Richler, Cheung, & Gauthier, 2011a). In the Feature Identity Task, I report that participants are more likely to respond same when status of the other feature was the 'same' than when 'different', replicating the results of Richler et al. (2008).

The replication of these standard behavioural results is important as it might be thought that the use of complete factorial designs changes the basic attributes of face processing. For example, the use of the complete factorial design requires attention to be distributed across both features. Richler et al. (2008) have shown the results are comparable for the composite face effect, but must be interpreted in a different way: as an effect of status of the other feature. The important point to note is that the use of complete factorial designs has not changed the signature findings associated with face processing, inversion and congruency.

With specific reference to the GRT violations, I focussed on the inferences made from the probit analysis (DeCarlo, 2003). Using probit analysis as the basis for drawing inferences within GRT for face processing is novel, although the methods for doing so rests in the work of Menneer and colleagues (Menneer, Silbert, Cornes, Wenger, Townsend, & Donnelly, 2009; Menneer, Wenger, & Blaha, 2010, 2012).

Evidence showed that for typical adults, the Feature Size Task does not lead to orientation specific violations and only generates reliable evidence of violations of perceptual independence (PI). The Feature Identity Task mostly leads to violations of decisional separability (DS), present in both upright and inverted versions of the task. The Feature Orientation Task, when considered across upright and inverted versions of the task, leads to violations of all three GRT constructs. Violations of perceptual separability (PS) were found for upright faces, and violations of DS in inverted faces. Violations of PI were found across both orientations. Consistent with previous modelling evidence (Cornes et al., 2011), these results show that evidence for configurality in the Thatcher illusion (Thompson, 1980) is not generated by a violation of PI only.

Analysis with the probit model revealed violations of PI across all three tasks, whilst the marginal method did not, suggesting the marginal method was insensitive to violations of PI. This has implications for previous studies that failed to find evidence for violations of PI using marginal analysis (e.g. Wenger & Ingvalson, 2002, 2003; Richler et al., 2008; Cornes et al., 2011).

The probit model has been shown to be particularly advantageous in understanding dependencies between features using mathematically specified definitions, but together the methods provide converging evidence for dependencies in the Thatcher illusion. It is clear that both methods need continued development and a formalised protocol to interpret the two outputs together if they contradict. The results of Chapter 2 have shown that GRT analysis is sufficiently sensitive to violations of PI in face processing tasks. This might now make the framework more attractive for others to explore GRT in face processing.

Chapter 3

In Chapter 3, I sought converging evidence for the sources of configurality in the Thatcher illusion revealed by the GRT analysis in Chapter 2 by using an event-related potential (ERP) approach. There were no face specific effects at the P1 component, consistent with Rouselett, Husk, Bennett and Sekuler (2008) who found no face effects at P1 when low level stimulus factors were controlled. There was a reduction in the amplitude of the N170 with level of Thatcherisation in the right hemisphere and inversion effects at the N170, P2 and P3b replicating previous findings. Amplitude was increased and delayed for inverted faces at the N170 (e.g. Boutsen, Humphreys, Praamstra, & Warbrick, 2006), increased for upright faces at P2 and increased for inverted faces at P3b. Effects at N170 (e.g. Rossion & Jacques, 2008) and P2 (e.g. Milivojevic, Clapp, Johnson, & Corballis, 2003) were localised to the right occipito-temporal area, while effects at P3b were strongest at the midline centro-parietal area (e.g. Polich, 2012), replicating previous findings. The effects of inversion were independent of the effect of Thatcherisation.

There is a large amount of variation between ERP studies exploring face processing. Typically ERP studies have focussed on the N170 as a perceptual marker of configurality, focusing on regions associated with perceptual processes. However, the study in Chapter 3 used a data driven approach to ERP analysis, highlighting areas associated with decisional processing. The results show a correspondence between GRT violations and ERP findings. One reading allows the suggestion that the Thatcherisation effect at N170 reflects a violation of PI, the inversion effect across N170 and P2 reflect violation of PS and the inversion effect at P3b reflects a violation of DS. Although mapping between violations in a GRT task to results of an ERP task has been attempted before (Keufner, Jacques, Alonso Prieto, & Rossion, 2010), this study was able to show evidence for potential decisional effects in the ERP. If these results were replicated, and the interpretation confirmed, then using an ERP marker to identify the potential of particular violations would be a useful tool for examining evidence for configural processing. Using multiple methods can provide sources of converging or disproving evidence, so it is important to pursue this approach in the future.

Chapter 4

In Chapter 4, I explored the utility and value of applying GRT to the issue of understanding the development of configural face processing. The issue was chosen because re-analysis of de Heering, Houthuys and Rossion (2007) and Mondloch, Pathman, Maurer, Le Grand and Schonen (2007) demonstrated underlying shifts in bias with development that had been previously unexplored. Conducting a complete factorial study with young children is challenging because of the number of trials involved and the potential for children to give up or provide unreliable data. For this first effort, data was collated across children such that GRT analysis was computed on aggregated data while standard analyses of sensitivity and bias was conducted at the individual level. Importantly, and as in Chapter 2, the standard analysis of sensitivity showed it improved with age, and interacted with orientation to show higher sensitivity in upright faces. In this regard the sensitivity data replicated previous developmental findings of gradual improvement in face processing with age (e.g. Donnelly & Hadwin, 2003). This is important as, like in Chapter 2, the use of a complete factorial design has not changed the fundamental nature of task performance in participants.

It is acknowledged that there is an issue with the data in Chapter 4. The violations reported for adults performing the Feature Orientation Task in Chapter 2 were not replicated. This is problematic as either many trials per participant is required to have enough power to produce similar violations or data is required from many more participants. Whichever of these alternatives is the case, the experiment reported in Chapter 4 is underpowered. Given that the experiment is underpowered caution should be taken in accepting the violation of PI found for 8/9 year olds or the failure to find any other violations.

There is one finding that suggests this issue is worthy of further experimentation. The inversion effect found in the probit correlations for adults is indicative of a qualitative change from children to adults. Nevertheless, further experimentation is required.

Chapter 5

The final experimental chapter examined the Thatcher illusion in a series of studies with a prosopagnosic individual (PHD). Unlike other tasks, there have been relatively few studies of the Thatcher illusion in prosopagnosia (Boutsen & Humphreys, 2002; Carbon, Grüter, Weber, & Lueschow, 2007). Therefore, Chapter 5 started by characterising PHDs performance in terms of a range of standard tasks, before exploring his deficits using a set of complete factorial studies. Ultimately, there was no evidence that PHD demonstrated any configural processing: a finding highlighted in the standard tasks. In these tasks, he was unable to categorise Thatcherised faces from typical faces, but was able to discriminate between them. He also had difficulties discriminating between some emotions consistent with an absence of configural processing. This supports the specific role of configurality in some emotion processing and is consistent with previous evidence from typical individuals (e.g. Calder, You, Keane, & Dean, 2000). However, in this study the relationship was demonstrated directly. The finding that PHD is not sensitive to the illusion and has emotional expression deficits is consistent with the findings of Donnelly et al. (2011) who have shown the Thatcher illusion involves activation of emotional cortices.

PHD was unable to perform in a meaningful way in the complete factorial GRT studies. This is probably a result of his profound difficulties in short-term visual memory. Although the GRT framework now seems unsuitable to assess sources of the configural deficit for PHD, this may not be true for other prosopagnosic patients. Along with the results of Chapter 4, it indicates that applying complete factorial designs (and therefore GRT analysis) in special populations is a non-trivial problem.

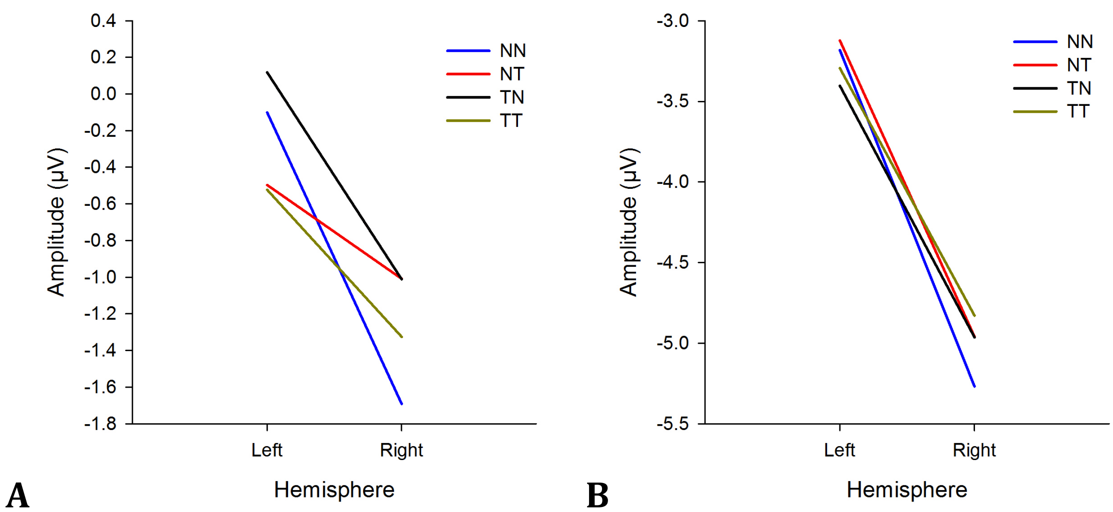


Figure 6.1. Graph showing (A) PHD and (B) controls split by hemisphere and condition for amplitude (µV) of the N170 component. The interaction is significant for the controls.

The results of Chapter 3, where the ERP markers seem to correspond to specific GRT violations, may provide a route through the impasse of using a complete factorial design with PHD. In a pilot study, the ERP responses of PHD to upright and inverted Thatcherised stimuli in the task used in Experiment 3.1 were recorded and plotted against the controls who took part in Experiment 3.1. These data are shown in Figure 6.1 and Figure 6.2. Although the data is just descriptive at this stage, the plot comparing hemisphere with Thatcherisation (Figure 6.1) seems to confirm no influence of Thatcherisation for PHD on the N170 in the right hemisphere like that observed for controls. Examining the ERP trace across the epoch when plotted by orientation condition, it is clear amplitudes are attenuated for PHD in comparison to controls. However, there appears to be an effect of orientation on both the N170 and

P2 (Figure 6.2a). There also appears to be evidence for inversion at P3b, which was very delayed in comparison to controls (Figure 6.2b). This may be evidence that PHD does have a different sensitivity for upright versus inverted faces, and a different bias to upright versus inverted faces: in GRT terms, violations of PS and DS but not PI. This approach certainly warrants further investigation.

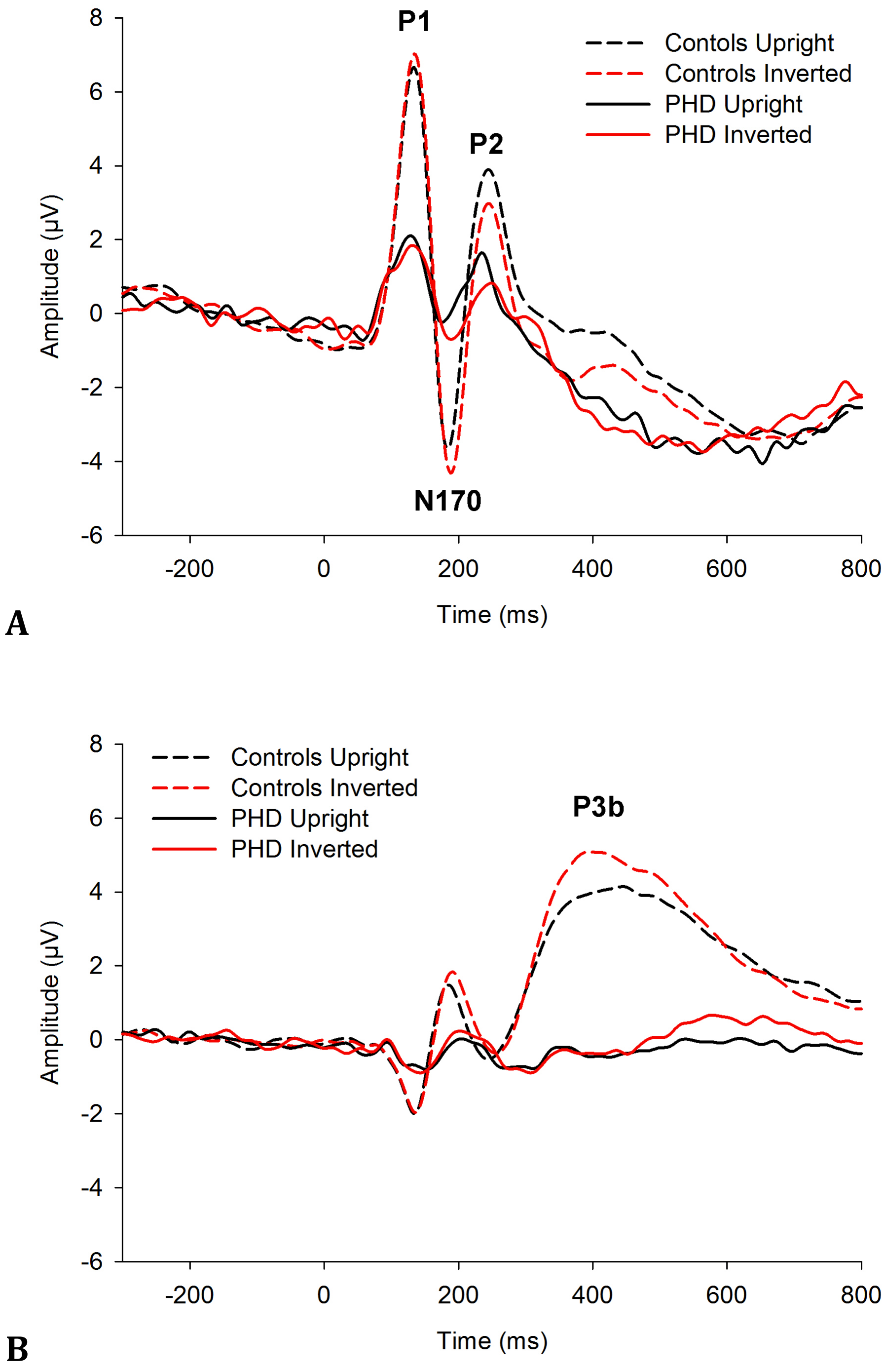


Figure 6.2. Amplitude (µV) across upright and inverted face conditions for PHD and controls from Experiment 3.1 at (A) the PO_8 electrode and (B) the CP_z electrode.

Conclusions and Future Directions

The research presented in this thesis shows there is a need to examine multiple sources of configurality in face processing. To do so requires the use of a robust theoretical framework that maps directly and unambiguously to testable hypotheses. Using GRT as the theoretical underpinning framework, I have demonstrated that by reframing common face processing tasks using a complete factorial design, basic underlying behavioural effects are retained. When analysed using GRT, performance in tasks requiring judgements of feature size, identity and orientation is underpinned by varying sources of configurality. In exploring the Feature Orientation Task further, I have shown that GRT violations might be reflected in ERP markers. Finally, there are challenges but also opportunities in applying GRT to specific populations, as revealed by studies with developmental and neuropsychological populations reported in this thesis.

The experiments in this thesis have provided some answers to important questions. However, they have led to many more questions of which I note three of these here. Firstly, in some conditions, violations were commonly but not always found. The issue is whether this is a result of statistical power or probability, or due to real individual differences in the sources of configurality used across tasks. Of course, in standard experiments, averaging across groups removes such individual differences. Recently some researcher have sought to explore the impact of individual differences on evidence for face processing across ERP (Rousselet et al. 2008) and information-processing signatures (Fifíc &Townsend,2010). Should it be found that there are reliable individual differences in sources of configurality found for specific tasks or conditions, it might suggest that sub-groups of participants exist. If so, it is possible we would be able to predict these sub-groups through differences in age, cognitive style, psychopathology or neuropsychology. Evidence for differences in cognitive style is consistent with local and global configural processing (Boutsen & Humphreys, 2003). Individuals with autism are reported to focus on local components (Behrmann, Avidan, Leonard, Kimchi, Luna, Humphreys, & Minshew, 2006), and the existence for different types of prosopagnosia, each showing a different profile of deficits (e.g. Goldsmith & Liu), suggests different sources of configurality have been affected.

The second issue is the status of converging evidence. In the present thesis, converging evidence was sought in terms of measuring ERP waveforms. In fact, the evidence is suggestive of a correspondence between GRT constructs and ERP waveforms. However, of interest would be a functional brain imaging study. Given evidence for perceptual and decisional sources of configurality in the Thatcher illusion, fMRI could be used to confirm whether there are two loci of effects for enhanced or reduced processing in relation to Thatcherised stimuli.

The final issue is that of familiarity. Humans are very good at recognising familiar people, even under difficult conditions. In contrast, recognition of unfamiliar people is relatively poor (Burton & Jenkins, 2011), with unfamiliar face recognition impaired by changes in pose, expression and context (Johnston & Edmonds, 2009). Cognitive neuropsychological case studies report dissociation between familiar and unfamiliar face recognition (Malone, Morris, Kay, & Levin, 1982), yet literature suggests the fundamental differences between familiar and unfamiliar face processing are not yet understood (Burton & Jenkins, 2011). The present experiments have focussed on unfamiliar face processing. How evidence for multiple sources of configurality fit into the distinction between familiar and unfamiliar face processing remains to be explored.

This thesis has provided an experimental and novel approach to investigating sources of configurality. A rigorous framework has been adopted, and a series of face processing tasks have been investigated to reveal they differ in sources of configurality. Findings in this thesis raise further questions in wider fields of study, which must now be examined further.

Appendices

Appendix A. The Value of Signal Detection Theory

The type of analysis performed on data from previous face processing tasks must be considered. Early research focused on accuracy, error rates and RTs, and configural processing was assumed to be perceptual. There was a lack of direct mapping of the analysis to theory. Furthermore, these studies rarely consider the role of changes in bias across conditions in generating measures for analysis. For identification and recognition of a stimulus, the brain must decide whether the information represents a target or a non-target. Signal detection theory (SDT, Green & Swets, 1966) is advantageous as it can separate decisional from perceptual effects. Macmillan and Creelman (2005) explain that targets and non-targets can be quite similar, so there is often a lot of noise in the perceptual process making identification of a target imperfect. SDT assumes that a participant will sample information from a trial in an experiment. This information will be represented numerically on a continuum, which is then used to make a decision about whether the stimulus is a target (signal) or not (noise). Signal and noise stimuli can be plotted on a continuum as two normal distributions which may overlap. Rather than using pure accuracy data, SDT uses sensitivity and bias based on the hit rates and false alarms in trials.

In SDT, the theoretical model comprises of the two distributions (signal and noise) and a decisional criterion. Sensitivity to the true state of the stimulus is measured by the degree to which the two distributions are separated (d'). The criterion is set as a cut point between making a decision about the stimulus as a signal or as noise. The location of the criterion determines the extent to which the observer will be biased to give a particular response (c). A schematic of this model can be seen in Figure A.

SDT provides a way for looking at face processing research and the way in which people make responses to stimuli. Since Green and Swets published their paper, the content of detection theory and the domains to which it is applied have broadened. Rather than just detecting a signal from noise, there are now models for different types of experimental designs (e.g. yes-no paradigm or 2AFC, Macmillan & Creelman, 2005).

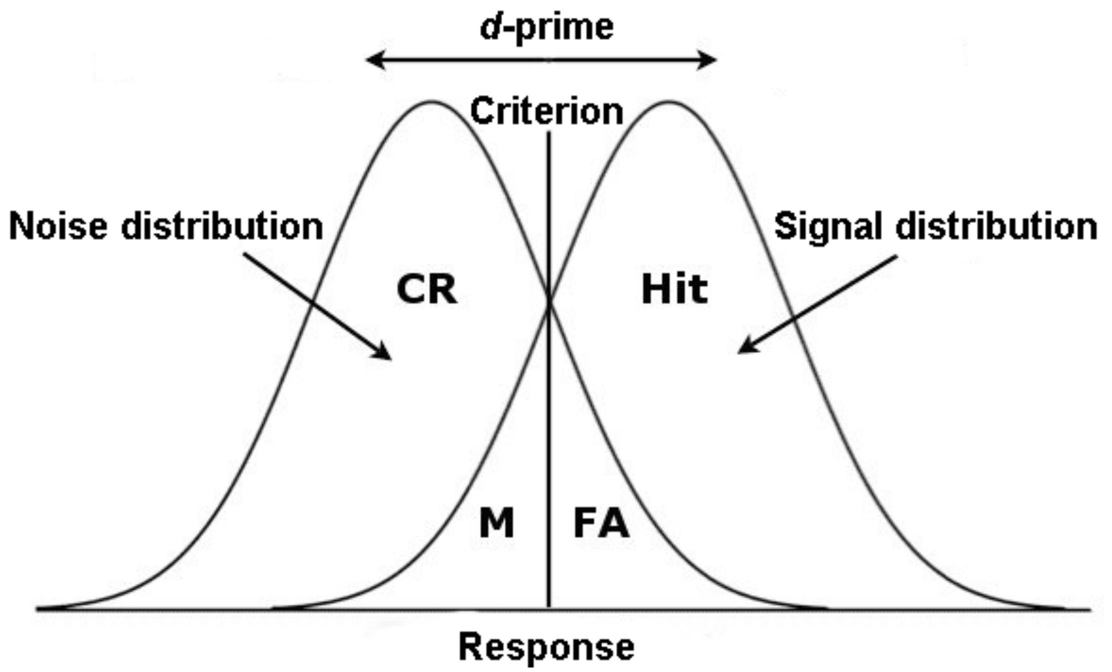


Diagram showing the signal-noise distribution, the four response outcomes and the d-prime and criterion measures. Labels represent hit, CR = correct rejection, M = miss, FA = false alarm.

Farah et al. (1998) provided hit rates in their original analysis of a selective attention face task (Experiment 1). Wenger and Ingvalson (2002) re-inspected their results using SDT and found that when irrelevant features differed, observers adopted a conservative response strategy compared to when they were the same. A shift in response criterion suggested decisional processes explained the pattern of results. This finding was controversial as configural processing has previously been assumed as a purely perceptual skill. This analysis provided an insight into the data that was not previously exploited, suggesting a need to start looking at face processing in this way (including consideration of response biases). The findings also enforced the need to re-examine the holistic perception hypothesis and the underlying assumptions using a formal testing method.

Signal detection analysis can only be performed on one dimension. As simultaneous responses to multiple stimulus dimensions are required in order to address the underlying assumptions of holism, a multidimensional methodology is needed to resolve the definition of configural processing.

Appendix B. Simulations for Known Violations of PI, PS and DS

Proportion of 1000 Confusion Matrices in which Violations were Detected by Marginal and Probit Analyses in Simulated Data with Known Violations.

Known violation	Analysis: (M)arginals or (P)robits	Violations detected (proportion) PI in (1,1)	PI in (2,1)	PI in (1,2)	PI in (2,2)	PS	DS
None	M	0.000	0.000	0.000	0.000	0.028	0.028
	P	0.019	0.016	0.009	0.013	0.002	0.000
PI in (1,1) Correlation = -0.25	M	**0.000**	0.000	0.000	0.000	0.025	0.022
	P	0.226	0.023	0.014	0.025	0.001	0.000
PI in (1,1) Correlation = +0.25	M	**0.000**	0.000	0.000	0.000	0.024	0.025
	P	0.305	0.024	0.027	0.027	0.002	0.000
PI in (1,1) Correlation = -0.5	M	**0.000**	0.000	0.000	0.000	0.031	0.022
	P	0.802	0.016	0.019	0.016	0.002	0.000
PI in (1,1) Correlation = +0.5	M	**0.000**	0.000	0.000	0.000	0.024	0.029
	P	0.929	0.013	0.015	0.027	0.002	0.000
PI in all distributions: Correlation = -0.25	M	**0.000**	**0.000**	**0.000**	**0.000**	0.023	0.017
	P	0.236	0.301	0.281	0.230	0.002	0.000
PI in all distributions: Correlation = +0.25	M	**0.000**	**0.000**	**0.000**	**0.000**	0.026	0.021
	P	0.309	0.208	0.238	0.274	0.004	0.000
PI in all distributions: Correlation = -0.5	M	**0.000**	**0.000**	**0.000**	**0.000**	0.019	0.022
	P	0.744	0.938	0.934	0.709	0.001	0.000
PI in all distributions: Correlation = +0.5	M	**0.000**	**0.000**	**0.000**	**0.000**	0.027	0.028
	P	0.927	0.726	0.774	0.935	0.002	0.000
PS: d' = 2 versus d' = 2.5	M	0.000	0.000	0.000	0.000	0.578	0.234
	P	0.020	0.014	0.023	0.008	0.347	0.019
PS: d' = 2 versus d' = 3	M	0.000	0.000	0.000	0.000	0.984	0.016
	P	0.015	0.012	0.011	0.030	0.964	0.099
DS with a continuous decision bound: c = 0 versus c = 0.25	M	0.000	0.000	0.000	0.000	0.021	0.609
	P	**0.240**	**0.255**	**0.250**	**0.349**	0.001	0.138
DS with a continuous decision bound: c = 0 versus c = 0.5	M	0.000	0.000	0.000	0.000	0.024	0.973
	P	**0.771**	**0.765**	**0.762**	**0.984**	0.000	0.944
DS with a piecewise decision bound: c = 0 versus c = 0.25	M	0.000	0.000	0.000	0.000	0.027	0.621
	P	0.020	0.014	0.018	0.018	0.000	0.144
DS with a piecewise decision bound: c = 0 versus c = 0.5	M	0.000	0.000	0.000	0.000	0.024	0.976
	P	0.013	0.009	0.009	0.011	0.001	0.942

Note. Due to space limitations, results are reported for single violations and for d' = 2 only, although results for d' = 1 are similar. Inferential errors are shown in bold. Perceptual independence (PI); perceptual separability (PS); decisional separability (DS).

Appendix C. A Basic Introduction to the ERP Technique

Measuring electrical activity in the human brain by placing electrodes on the scalp, amplifying the signal and plotting the change in voltage over time is called electrocephalogram (EEG). EEG has proven to be useful in many aspects of science. However, EEG is a coarse measure of brain activity as there are many contributing sources of activity making it difficult to isolate any individual processes. Within the EEG are neural responses associated with specific events. It is possible to extract these responses from the overall EEG to obtain event-related potentials (ERPs). The ERP technique can be exploited to study brain activity in relation to cognition.

Important aspects of electricity and magnetism must be considered for studying neural origins of ERP. Three components are fundamental to understanding electricity: voltage, current and resistance.

Voltage is electrical pressure, also called potential as it is the potential for electricity to flow from one place to another. It is measured in volts and usually labelled as V. Current is the number of charged particles (electrons) that flow past a given point in a specific amount of time. Current is measured in amperes, abbreviated to amps and is usually labelled I for intensity. Resistance is the ability of a substance to keep charged particles from passing, the opposite of conductance, it is measured in ohms (Ω) and is usually labelled R. The relationship between these three components is summed up by Ohm's Law, where each component can be calculated from a simple formula if the other two values are known. To use an example of water, if a hose pipe is very long and thin (greater resistance), a high pressure (large voltage) will be required to fill a bucket of water in a short amount of time (high current).

These three principles relate to the ERP technique as EEG is recorded as a potential for current to pass between two electrodes. This is important as it means recordings can never be made from a single scalp electrode. Noise in the EEG recording is any source of variation in the data that is unrelated to what the experimenter is trying to record. Noise can come from a range of sources including tension, eye movement, sweat on the skin and electrical equipment in the recording environment. Experimenters want to know the level of noise in the recording; therefore they should use a common reference point. To solve the problem of a ground electrode also picking up noise, three electrodes are used, an active electrode, a reference electrode and a ground electrode.

There are also many other factors such as medications, lack of sleep, caffeine and smoking that can affect the quality of an EEG recording, therefore a pre-screening questionnaire and strict exclusion criteria are often used. Also, in order to reduce the sources of noise in the recording itself, many filtering techniques can be used to remove particular frequencies which are associated with different types of noise.

Impedance is the appropriate name for resistance when current varies over time. As ERPs vary over time, impedance is the most relevant concept. In EEG recording, experimenters want to keep impedance (the resistance between electrodes and scalp) as low as possible. Therefore, in the preparation of the electrodes the experimenters exfoliate the skin site where the recording will be taken and clean away the layer of dry skin cells responsible for higher impedance values. It is common practice to keep impedance below 5 KΩ. The sampling rate is the number of samples taken per second (e.g. 500Hz is a sample every 2ms), the higher the sampling rate, the greater the temporal precision. Higher sampling rates are required for early sensory responses.

Compared to imaging studies with high spatial resolutions, low spatial resolution in ERP studies means it is difficult to localise an effect at a component to a particular electrode site, therefore it is often more informative to investigate an electrode cluster. However, ERP has fantastic temporal resolution so can separate different temporal components and how these may be influenced by levels of stimuli manipulation as modulation of ERPs can occur very early after stimulus onset.

The number of trials is usually must larger than required as many will be excluded due to artefact. EEG/ERP studies must be designed with care as longer studies lead to poorer quality data as participants will find it hard to stay still and comfortable. However, more trials can be recorded which allows for a greater number of trials per condition after pre-processing. Pre-processing is the stage before analysis where data is filtered and cleaned to remove artefacts through excluding particular frequencies and correcting for eye movements. Incorrect and noisy trials are rejected from analysis bad electrodes with poor impedance values are removed.

Component letters identify whether it is a positive or negative voltage. The number either refers to the order of component or the time at which the component occurs in milliseconds: E.g. P1 is the first positive peak which usually occurs at about 100 ms so is often called P100. In some tasks, latency of a component can be very specific so is often referred to with a millisecond name when it is investigated in a particular field, e.g. the N170 which is the name for the N1 component when examining face perception as it usually peaks at approximately 170ms.

There are many inconsistencies in how ERP studies are presented. Plotting and ERP waveform can either be done with positive up or negative up. There is a lot of variation in the literature and across software but either form is accepted. Also, different methods of identifying latency windows for analysis and identifying electrodes to examine are employed. Studies also have different cap set ups so the electrode locations vary. I used the 10-20 system as this is the typical set up for most ERP studies in the face processing literature and allows comparison to effects in previous studies as the electrode locations are the same. However, there are still some inconsistencies in naming electrodes, the P_7 and P_8 locations are equivalent to T_5 and T_6 in the old nomenclature (Privik, Broughton, Coppola, Davidson, Fox, & Nuwer, 1993) and often not all electrode sites will be used in the EEG recording.

Bibliography

Luck, S. J. (2005). *An introduction to the event-related potential technique.* Cambridge, MA: MIT Press.

Luck, S. J., & Kappenman, E. S. (2012). *The Oxford handbook of event-related potential components.* New York, NY: Oxford University Press.

Pivik, R. T., Broughton, R. J., Coppola, R., Davidson, R. J., Fox, N., & Nuwer, M. R. (1993). Guidelines for the recording and quantitative analysis of electroencephalographic activity in research contexts. *Psychophysiology 30,* 547– 558.

Appendix D. Pre-Screening Questionnaire for ERP Study

Contact phone number: ……………………………

Date of birth: ……………………………………

Sex (circle): **Male / Female**

Your handedness (circle): **Left / Right**

Ethnicity: ………………………………………………………………………………………...........

Please circle Yes or No to the following questions:

Have you consumed any product containing caffeine in the last two hours? **Yes/No**

Do you smoke **Yes/No**

If yes please specify how many cigarettes you have smoked today prior to the experiment:

Are you currently on any form of medication? **Yes/No**

If yes please specify:……………………………………………………………………………

Have you ever suffered an epileptic seizure? **Yes/No**

If yes please specify:……………………………………………………………………………

Have you ever suffered any serious head injuries or periods of unconsciousness? **Yes/No**

If yes please specify:……………………………………………………………………………

Do you have any vision problems? **Yes/No**

If yes please specify:……………………………………………………………………………

Do you have any hearing problems? **Yes/No**

If yes please specify:……………………………………………………………………………

Is English your first language? **Yes/No**

If no, how many years have you been a fluent English speaker? ____________

Have you used any psychoactive substance in the past 24 hours? **Yes/No**

Have you used any psychoactive substance more than once a month in the last 6 months? **Yes/No**

How many hours do you sleep on average every night? ____________

How many hours sleep did you have last night? ____________

References

Adams, W. J., Gray, K. L. H., Garner, M., & Graf, E. W. (2010). High-level face adaptation without awareness. *Psychological Science, 21,* 205-210.

Allison, T., Puce, A., Spencer, D. D., & McCarthy, G., (1999). Electrophysiological studies of human face perception. I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. *Cerebral Cortex, 9,* 415-430.

Ashby, F. G., & Townsend, J. T. (1986). Varieties of perceptual independence. *Psychological Review, 93,* 154-179.

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

Batty, M., & Taylor, M. J. (2003). Early processing of the six basic facial emotional expressions. *Cognitive Brain Research, 17,* 613–620.

Behrmann, M., Avidan, G., Leonard, G. L., Kimchi, R., Luna, B., Humphreys, K., & Minshew, N. (2006). Configural processing in autism and its relationship to face processing. *Neuropsychologia, 44,* 110–129.

Bentin, E., & Bhatt, R. S. (2004). The Thatcher illusion and face processing in infancy. *Science, 7,* 431-436.

Bentin, S., Allison, T., Puce, A., Perez, E., & McCarthy, G. (1996). Electrophysiological studies of face perception in humans. *Journal of Cognitive Neuroscience, 8,* 551–565.

Bentin, S., Deouell, L. Y., & Soroker, N. (1999). Selective visual streaming in face recognition: evidence from developmental prosopagnosia. *NeuroReport, 10,* 823–827.

Benton, A. L., Sivan, A. B., Hamsher, K. de S., Varney, N. R., Spreen, O. (1994). *Contributions to neuropsychological assessment (2nd ed.).* New York: Oxford University Press.

Bodamer, J. (1947). Die prosopagnosie. *Archiv für Psychiatrie und ervenkrankheiten, 179,* 6-54. Partial English translation by Ellis, H. D., & Florence, M. (1990). *Cognitive Neuropsychology, 7,* 81-105.

Boutsen, L., & Humphreys, G. W. (2002). Face context interferes with local part processing in a prosopagnosic patient. *Neuropsychologia, 40,* 2305-2313.

Boutsen, L., & Humphreys, G. W. (2003). The effect of inversion on the encoding of normal and “thatcherized” faces. *The Quarterly Journal of Experimental Psychology, 56,* 955-975.

Boutsen, L., Humphreys, G. W., Praamstra, P., & Warbrick, T. (2006). Comparing neural correlates of configural processing in faces and objects: An ERP study of the Thatcher illusion. *NeuroImage, 32,* 352-367.

Burton, A. M., & Jenkins, R. (2011). Unfamiliar face perception. In A. J. Calder, G. Rhodes, M. H. Johnson & J. V. Haxby (Eds), *Handbook of face perception* (pp. 287-306). New York: Oxford University Press.

Calder, A. J., Young, A. W., Keane, J., & Dean, M. (2000). Configural information in facial perception. *Journal of Experimental Psychology, 26,* 527-551.

Carbon, C. C., Grüter, T., Weber, J. E., & Lueschow, A. (2007). Faces as objects of non-expertise: Processing of Thatcherised faces in congenital prosopagnosia. *Perception, 36,* 1635-1645.

Carbon, C. C., Schweinberger, S. R., Kaufman, J. M., Leder, H. (2005). The Thatcher illusion seen by the brain: An event-related brain potential study. *Cognitive Brain Research, 24,* 544-555.

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 configuration more by adults than by children? *Visual Cognition, 1,* 254-274.

Cheung, O. S., & Gauthier, I. (2010). Selective interference on the holistic processing of faces. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 448–461.

Cornes, K., Donnelly, N., Godwin, H. J., & Wenger, M. J. (2011). Perceptual and decisional factors influencing the discrimination of inversion in the Thatcher illusion. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 645–668.

Crawford, J. R., Garthwaite, P. H., & Porter, S. (2010). Point and interval estimates of effect sizes for the case‑controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cognitive Neuropsychology, 27,* 245-260.

Crawford, J. R. & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12,* 482-486.

Crookes, K., & McKone, E. (2009). Early maturity of face recognition: No childhood development of holistic processing, novel face encoding, or face-space. *Cognition, 111,* 219-247.

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

de Gelder, B., Frissen, I., Barton, J., & Hadjikhani, N. (2003). A modulatory role for facial expressions in prosopagnosia. *Proceedings of the National Academy of Sciences, 100*, 13105-13110.

de Gelder, B., & Van den Stock, J. (2011). The bodily expressive action stimulus test (BEAST). Constructon and validation of a stimulus basis for measuring perception of whole body expression of emotions. *Frontiers in Psychology, 2,* 181.

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology, 96,* 57-70.

DeCarlo, L. T. (1998). Signal detection theory and generalized linear models. *Psychological Methods, 3,* 186-205.

DeCarlo, L. T. (2003). Source monitoring and multivariate signal detection theory, with a model for selection. *Journal of Mathematical Psychology, 47,* 292-303.

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

Donnelly, N., Cornes, K., & Menneer, T. (2012). An examination of the processing capacity of features in the Thatcher illusion. *Attention, Perception & Psychophysics, 74,* 1475-1487.

Donnelly, N., & Davidoff, J. (1999). The mental representations of faces and houses: Issues concerning parts and wholes. *Visual Cognition, 6,* 319-343.

Donnelly, N., & Hadwin, J. A. (2003). Children's perception of the Thatcher illusion: Evidence for development in configural face processing. *Visual Cognition, 10,* 1001-1017.

Donnelly, N., Zürcher, N. R., Cornes, K., Snyder, J., Naik, P., Hadwin, J. A., & Hadjikani, N. (2011). Discriminating grotesque from typical faces: Evidence from the Thatcher illusion. *PLoS ONE, 6,* e23340.

Duchaine, B. C., Parker, H., & Nakayama, K. (2003). Normal recognition of emotion in a prosopagnosic. *Perception*, 32, 827-838.

Eimer, M. (2000a). Event-related brain potentials distinguish processing stages involved in face perception and recognition. *Clinical Neuropsychology, 111,* 694-705.

Eimer, M. (2000b). Effects of face inversion on the structural encoding and recognition of faces: Evidence from event-related brain potentials. *Cognitive Brain Research, 10,* 145-158.

Eimer, M., & McCarthy, R. A. (1999). Prosopagnosia and structural encoding of faces: Evidence from event-related potentials. *NeuroReport, 10,* 255-259.

Ekman, P. & Friesen, W. V. (1976). *Pictures of facial affect.* Palo Alto, CA: Consulting Psychology Press.

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. N. (1995). 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.

Fifíc, M., & Townsend, J. T. (2010). Information-processing alternatives to holistic perception: identifying the mechanisms of secondary-level holism within a categorization paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36,* 1290-1313.

Fonaryova Key, A. P., Dova, G. O., & Maguire, M. J. (2005). Linking brainwaves to the brain: An ERP primer. *Developmental Neuropsychology, 27,* 183-215.

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

George, N., Jemel, B., Fiori, N., Chaby, L., & Renault, B. (2005). Electropsysiological correlates of facial decision: insights from upright and upside-down Mooney face perception. *Cognitive Brain Research, 24,* 663-673.

Goffaux, V., & Rossion, B. (2006). Faces are “spatial”- Holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 1023-1039.

Goldsmith, Z. G., & Liu, G. T. (2001). Facial recognition and prosopagnosia: Past and present concepts. *Neuro-Ophthalmology, 25,* 177-192.

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

Gu, X., Li, Y., Yang, X., & Zhu, Y. (2007). Spatial-temporal analysis of face processing using an ERP study of the Thatcher illusion. *Proceedings of the 29th Annual International Conference of the IEEE EMBS,* 2496-2499.

Halit, H., de Haan, M., & Johnson, M. H. (2000). Modulation of event-related potentials by prototypical and atypical faces. *Neuroreport, 11,* 1871-1875.

Hancock, P. J. B., Bruce, V., & Burton, A. M. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Psychology, 4,* 330-337.

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

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,* 189-199.

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

Ingvalson, E. M., & Wenger, M. J. (2005). A strong test of the dual-mode hypothesis. *Perception and Psychophysics, 67,* 14-35.

Itier, R. J., & Taylor, M. J. (2002). Inversion and contrast polarity reversal affect both encoding and recognition processes of unfamiliar faces: A repetition study using ERPs . *NeuroImage, 15,* 353-372.

Itier, R. J., & Taylor, M. J. (2004). N170 or N1? Spatiotemporal differences between object and face processing using ERPs. *Cerebral Cortex, 14,* 132-142.

Jacques, C., & Rossion, B. (2007). Early electrophysiological responses to multiple face orientations correlate with individual discrimination performance in humans. *NeuroImage, 36,* 863-876.

Johnston, J. C., & McClelland, J. C. (1980). Experimental tests of a hierarchical model of word identification. *Journal of Verbal Learning and Verbal Behaviour, 19,* 503–524.

Johnston, R. A., & Edmonds, A. J. (2009). Familiar and unfamiliar face recognition: A review. *Memory, 17,* 577-596.

Kadlec, H. (1995). Multidimensional signal detection analysis (MSDA) for testing seperabitily and independence: A Pascal program. *Behavior Research Methods, Instruments, & Computers, 27,* 442–458.

Kadlec, H. (1999). MSDA2: Updated version of software for multidimensional signal detection analyses. *Behavior Research Methods, Instruments, & Computers, 31,* 384–385.

Kadlec, H., & Hicks, C. L. (1998). Invariance of perceptual spaces and perceptual separability of stimulus dimensions. *Journal of Experimental Psychology: Human Perception and Performance,* 24, 80-104.

Kadlec, H., & Townsend, J. T. (1992a). Implications of marginal and conditional detection parameters for the separabilities and independence of perceptual dimensions. *Journal of Mathematical Psychology, 36,* 325-374.

Kadlec, H., & Townsend, J. T. (1992b). Signal detection analysis of dimensional interactions. In F. G. Ashby (Ed.), *Multidimensional models of perception and cognition* (pp. 181–228). Hillsdale, NJ: Erlbaum.

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

Kuefner, D., Jacques, C., Alonso Prieto, E., & Rossion, B. (2010). Electrophysiological correlates of the composite face illusion: Disentangling perceptual and decisional components of holistic face processing in the human brain. *Brain and Cognition, 74,* 225–238.

Lehmann, D., & Skrandies, W. (1980). Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalography and Clinical Neurophysiology, 48,* 609-621.

Lewis, M. B. (2001). The Lady’s not for turning: Rotation of the Thatcher illusion. *Perception, 30,* 769-774.

Lewis, M. B. (2003). Thatcher's children: Development and the Thatcher illusion. *Perception, 32,* 1415-1421.

Macho, S. (2007). Feature sampling in detection: Implications for the measurement of perceptual independence. *Journal of Experimental Psychology: General, 136*, 133–153.

Mack, M. L., Richler, J. J., Gauthier, I., & Palmeri, T. (2011). Indecision on decisional separability, *Psychonomic Bulletin & Review, 18,* 1-9.

Macmillan, N. A. & Creelman, C. D. (2005). *Detection theory: A user's guide (2nd ed.).* New York: Lawrence Erlbaum Associates.

Macrae, C. N., & Lewis, H. L. (2002). Do I know you? Processing orientation and face recognition. *Psychological Science, 13,* 194-196.

Malone, D. R, Morris, H. H., Kay, M. C., & Levin, H. S. (1982). Prosopagnosia; a double dissociation between the recognition of familiar and unfamiliar faces. *Journal of Neurology, Neurosurgery & Psychiatry, 45,* 820-822.

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

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

McDonald, S., Flanagan, S., Rollins, J., Kinch, J. (2003). TASIT: A new clinical tool for assessing social perception after traumatic brain injury. *Journal of Head Trauma Rehabilitation. 18,* 219-38.

McKelvie, S. J. (1995). Emotional expression in upside-down faces: Evidence for configurational and componential processing. *British Journal of Sococial Psychology, 34,* 325-34.

McKenna, P., & Warrington, E.K. (1983). *Graded Naming Test.* Windsor, UK: NFER-Nelson.

Menneer, T., Silbert, N., Cornes, K., Wenger, M. J., Townsend, J. T. & Donnelly, N. (2009). Contrasting methods of model estimation for configural and holistic perception. *Journal of Vision, 9,* 555. Poster presented at the Vision Sciences Society Conference, Naples, FL).

Menneer, T., Wenger, M. J., & Blaha, L. (2010). Inferential challenges for general recognition theory: Mean-shift integrality and perceptual configurality.

Journal of Vision, 10, 1211. (Poster presented at Vision Sciences Society Conference, Naples, FL).

Menneer, T., Wenger, M. J., & Blaha, L. M. (2012). Probit analysis for multidimensional signal detection: An evaluation and comparison with standard analyses. Manuscript in preparation.

Mercure, E., Dick, F., & Johnson, M. K. (2008). Featural and configural processing differentially modulate ERP components. *Brain Research, 1239,* 162-170.

Mestry, N. (2008). *Evidence from prosopagnosia: Is N170 necessary for configural processing?* (Undergraduate Dissertation). University of Southampton, UK.

Mestry, N., Donnelly, N., Menneer, T., & McCarthy, R. A. (2012). Discriminating Thatcherised from typical faces in a case of prosopagnosia. *Neuropsychologia, 50,* 3410-3418.

Mestry, N., Menneer, T., Wenger, M. J., & Donnelly, N. (2012). Identifying sources of configurality in three face processing tasks. *Frontiers in Psychology, 3,* 456.

Milivojevic, B., Clapp, W. C., Johnson, B. W., & Croballis, M. C. (2003). Turn that frown upside down: ERP effects of Thatcherisation on misorientated faces. *Psychophysiology, 40,* 967-978.

Mondloch, C. J., Dobson, K. S., Parsons, J., & Maurer, D. (2004). Why 8-year-olds cannot tell the difference between Steve Martin and Paul Newman: factors contributing to the slow development of sensitivity to the spacing of facial features. *Journal of Experimental Child Psychology, 89,* 159-181.

Mondloch, C. J., Le Grand, R., & Maurer, D. (2002). Configural face processing develops more slowly than featural processing. *Perception, 31,* 553-566.

Mondloch, C. J., Pathman, T., Maurer, D., Le Grand, R., & de Schonen, S. (2007). The composite face effect in six-year-old children: Evidence for adult like holistic face processing. *Visual Cognition, 15,* 564-577.

Murray, J. E., Yong, E., & Rhodes, G. (2000). Revisiting the perception of upside down faces. *Psychological Science, 11,* 492-496.

O'Toole, A. J., Wenger, M. J., & Townsend, J. T. (2001). Quantitative models of perceiving and remembering faces: Precedents and possibilities. In M. J. Wenger & J. T. Townsend (Eds.) *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 1-38). Mahwah, NJ: Erlbaum.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologica, 9,* 97-113.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In J. Long & A. Baddeley (Eds.), *Attention and performance IX* (pp. 135–151). Hillsdale, NJ: Erlbaum.

Parks, T. E., Coss, R. G., & Coss, C. S. (1985). Thatcher and the Cheshire cat: Context and the processing of facial features. *Perception, 14,* 747 – 754.

Peelen, M. V., Lucas, N., Mayer, E., & Vuilleumier, P. (2009). Emotional attention in acquired prosopagnosia. *SCAN*, *4,* 268-277.

Pellicano, E., & Rhodes, G. (2003). Holistic processing of faces in preschool children and adults. *Psychological Science, 14,* 618-622.

Pellicano, E., Rhodes, G., & Peters, M. (2006). Are preschoolers sensitive to configural information in faces? *Developmental Science, 9,* 270-277.

Philiastides, M. G., & Sajda, P. (2006). Temporal characterization of the neural correlates of perceptual decision making in the human brain. *Cerebral Cortex, 16,* 509–518.

Picton, T. W., Bentin, S., Berg, P., Donchin, E., Hillyard, S. A., Johnson, J. R., ... Taylor, M. J. (2000). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37,* 127–152.

Polich, J. (2012). Neuropsychology of P300. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 159-207). New York, NY: Oxford University Press.

Reber, R., Winkielman, P., & Schwartz, N. (1998). Effects of perceptual fluency on affective judgements. *Psychological Science, 9,* 45-48.

Regan, D. (1989). *Human brain electrophysiology: Evoked potentials and evoked magnetic fields in science and medicine.* New York: Elsevier.

Rhodes, G., & Tremewan, T. (1996). Averageness, exaggeration, and facial attractiveness. *Psychological Science, 7,* 105-110.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011a). Holistic processing predicts face recognition. *Psychological Science, 22,* 464–471.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011b). Beliefs alter holistic face processing… if response bias is not taken into account. *Journal of Vision, 11,* 1–13.

Richler, J. J., Gauthier, I., Wenger, M. J., & Palmeri, T. J. (2008). Holistic processing of faces: Perceptual and decisional components. *Journal of Experimental Psychology: Learning, Memory and Cognition, 34,* 328-342.

Richler, J. J., Mack, M. L., Palmari, T. J., & Gauthier, I. (2011). Inverted faces are (eventually) processed holistically. *Vision Research, 51,*333–342.

Rock, I. (1973). *Orientation and form*. New York: Academic Press.

Rossion, B., Delvenne, J. F., Debatisse, D., Goffaux, V., Bruyer, R., Crommelinck, M., & Guérit, J. M. (1999). Spatio-temporal brain localization of the face inversion effect. *Biological Psychology, 50,* 173-189.

Rossion, B., Gauthier, I., Tarr, M. J., Despland, P. Bruyer, R., Linotte S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. *Neuroreport, 11,* 69-74.

Rossion, B., & Jacques, C. (2008). Does physical interstimulus variance account for early electrophysiological face sensitive responses in the human brain? Ten lessons on the N170. *NeuroImage, 39,* 1959–79.

Rossion, B., & Jacques, C. (2012). The N170: Understanding the time course of face perception in the human brain. In S. J. Luck & E. S. Kappenman (Eds.), *The Oxford handbook of event-related potential components* (pp. 115-141). New York: Oxford University Press.

Rothstein, P., Malach, R., Hadar, U., Graif, M., & Hendler, T. (2001). Feeling of features: different sensitivity to emotion in high-order visual cortex and amygdala. *Neuron, 32,* 747–757.

Rouse, H., Donnelly, D., Hadwin, J. A., & Brown, T. (2004). Do children with autism perceive second-order relational features? The case of the Thatcher illusion. *Journal of Child Psychiatry, 45,* 1246-1257.

Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2008). Time course and robustness of ERP object and face differences. *Journal of Vision, 8,* 1-18.

Scapinello, K. F., & Yarmey, A. D. (1970). The role of familiarity in immediate and delayed recognition of pictoral stimuli. *Psychonomic Science, 21,* 329-330.

Schupp, H. T., Öhman, A., Junghöfer, M., Weike, A. I., Stockburger, J., & Hamm, A. O. (2004). The facilitated processing of threatening faces: an ERP analysis. *Emotion, 4,* 189-200.

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

Semlitsch, H. V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts applied to the P300 ERP. *Psychophysiology, 23,* 695–703.

Shallice, T., (1990). *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.

Stephan, C. B. M., Breen, N., & Caine, D. (2006). The recognition of emotional expression in prosopagnosia: Decoding whole and face parts. *Journal of the International Neuropsychological Society*, *12,* 884-895.

Stürzel, F., & Spillmann, L. (2000). Thatcher illusion: Dependence on angle of rotation. *Perception, 29,* 937 – 942.

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

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

Tarr, M. J., & Pinker, S. (1989).Mental rotation and orientation dependence in shape recognition. *Cognitive Psychology, 21,* 233–282.

Thompson, P. (1980). Margaret Thatcher: A new illusion. *Perception, 9,* 483-484.

Tottenham, N., Tanaka, J. W., Leon, A. C., McCarry, T., Nurse, M., Hare, T. A., . . . Nelson, C. (2009). The NimStim set of facial expressions: Judgments from untrained research participants. *Psychiatry Research, 168,* 242–249.

Townsend, J. T., Hu, G. G., & Ashby, F. G. (1981). A test of visual feature sampling independence with orthogonal straight lines. *Psychological Research, 43,* 259–275.

Townsend, J. T., & Nozawa, G. (1995). Spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology, 39,* 321-360.

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., & Bruce, V. (1988). Mental rotation of faces. *Memory and Cognition, 16,* 556-566.

Verleger, R., Jaśkowski, P., & Wascher, E. (2005). Evidence for an integrative role of P3b in linking reaction to perception. *Journal of Psychophysiology, 19,* 165–181.

Warrington, E. K. (1984). *Recognition Memory Tests.* Berkshire, UK: NFER-Nelson.

Warrington, E. K. (1996). *Camden Memory Tests.* Hove, UK: Psychology Press.

Warrington, E. K., & James, M. (1991). *The Visual Object and Spacial Perception Battery.* Burry St. Edmunds, UK: Thames Valley Test Co.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—3rd Edition (WAIS-3®).* San Antonio, TX: Harcourt Assessment.

Wenger, M. J., & Ingvalson, E. M. (2002). A decisional component of holistic encoding. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28,* 872-892.

Wenger, M. J., & Ingvalson, E. M. (2003). Preserving informational separability and violating decisional separability. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 1106–1118.

Wenger, M. J., & Townsend, J. T. (2001). Faces as Gestalt stimuli: Process characteristics. In M. J. Wenger & J. T. Townsend (Eds.), *Computational, geometric, and process perspectives on facial cognition: Contexts and challenges* (pp. 229–284). Mahwah: Erlbaum.

Yarmey, A. D. (1971). Recognition memory for familiar "public" faces: Effects of orientation and delay. *Psychonomic Science, 24,* 286-288.

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

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

Young, A. W., Perrett, D. I., Calder, A. J., Sprengelmeyer, R., & Ekman, P. (2002). Facial expressions of emotion: Stimuli and tests (FEEST). (Bury St Edmunds Thames Valley Test Company, Eds.) *Psychology*, *126*, 420.