Cognitive Templates for Human Face Detection

A thesis submitted for the Degree of Ph.D. in the Division of Human and Social Sciences at the University of Kent

Alice Nevard

School of Psychology
University of Kent
September 2023

Abstract

Faces convey a range of important social information such as a person's identity, emotional state and direction of gaze. However, before such information can be attained from a face, it first has to be localised within the visual environment. Therefore, face *detection* is a key process that facilitates social navigation in our environment. Detection seems to be most sufficient when faces are presented in frontal, upright views and in colour with all features intact. Yet, the results of such research are mixed, due to the range of methods used. Therefore, this thesis explores the visual processes of face detection, by focusing on how faces are displayed, and the facial features utilised in finding faces in such displays. Firstly, the attentional draw of person stimuli in naturalistic scenes is considered, and the impact of the frequency of person presence on attentional draw (Chapter 2). The findings indicate that people draw attention spontaneously in naturalistic scenes, but at high rates of presentation such stimuli can be actively ignored in favour of other task demands.

The drivers of the detection process were then explored, by investigating the influence of display context, whilst noting the facial features that facilitate optimal detection (Chapter 3). The detection of frontal/upright faces and manipulated faces were comparable in blank displays, yet disparities between frontal/upright and manipulated faces emerge in array and scene displays, indicating that detection processes are dependent on scene complexity that requires search. Furthermore, the use of face manipulations revealed that external features, such as overall shape, drive detection.

Subsequently, the role of internal and external facial features in peripheral detection was considered using a novel gaze-contingent detection paradigm (Chapter 4). In the periphery, localisation is potentially enhanced for faces without internal features, with faster fixations on initially featureless faces. Yet performance declines when features are removed at fixation, suggesting they are essential for face classification, but not localisation. This was then reinforced, as faces with manipulated internal features (i.e. rotated) were difficult to distinguish from intact faces in the periphery. Therefore, internal features cannot be perceived clearly enough in the periphery to drive detection, but seem to gain importance at fixation when classifying faces.

Altogether, this thesis outlines the cognitive templates used in face detection that depend on presentation context. A colour-shape template based on skin tone and oval shapes seems to drive detection when localising faces. Yet a more detailed template with intact featural configuration is utilised closer to fixation once faces are being classified. Separating the processes involved in detection and the features involved in such processes helps to build a more coherent theory of face detection.

Acknowledgements

Firstly, I would like to thank my supervisor, Professor Markus Bindemann, for his advice, regardless of whether I listened to it. I have appreciated his guidance throughout this PhD and he has undoubtedly helped me develop into a better researcher and writer. I would also like to thank Dr Graham Hole, who encouraged me from my undergraduate to pursue this PhD and has supported me over the years.

I would like to thank my parents, close family and friends, who were always there whenever I needed them. Although they may not understand my work, they have supported me nonetheless. Oli, Jade and Jen deserve a special mention. To anyone who has been a part of this journey, thank you.

Finally, I would like to thank Peter, who has been my biggest supporter over the last 2 years. You have given me endless kindness and compassion through the best times and the tough times, and I appreciate everything you do for me. Of course, I could have done this without you, but it has certainly been better with you.

Declaration

I declare that this thesis is my own work carried out under the normal terms of supervision

A. Nevard

Alice Nevard

Publications

Chapter 3 (Experiments 4 to 8) of this thesis has been accepted for publication:

Nevard, A., Hole, G, J., Prunty, J., & Bindemann, M. (2023). Understanding face detection with visual arrays and real-world scenes. *Visual Cognition,* 31(5). https://doi.org/10.1080/13506285.2023.2277475

Chapter 4 (Experiments 9 to 12) of this thesis is in preparation for publication.

Table of contents

Chapter 1:

General Introduction

1.1 Introduction

Humans frequently encounter other people in their environment and this is important for a number of social processes, such as recognising a person's identity, their emotional state, and the direction of gaze to determine their attentional focus. For these processes to take place, people must first be located in the visual environment. This cognitive process of *detection* appears to proceed with ease, even under seemingly demanding conditions. A plethora of research has demonstrated, for example, that the human visual system is drawn towards peoples' faces when these are presented in visual displays (e.g., Langton, Law, Burton & Schweinberger, 2008; Ro, Russell & Lavie, 2001). This face advantage has been demonstrated in visual search tasks, in which human faces are found faster than non-human faces and other objects in stimulus arrays (e.g., Di Giorgio, Turati, Altoè & Simion, 2012; Maylott, Sansone, Jakobsen, & Simpson, 2021; Simpson, Maylott, Leonard, Lazo, & Jakobsen, 2019; Yang, Shih, Cheng, & Yeh, 2009). Faces are also preferentially looked at during free-viewing tasks, in which observers are not given specific task instructions. This demonstrates that these stimuli must hold significance, whether that be due to visual saliency or social importance (e.g., Birmingham, Bischof & Kingstonee, 2008). During search tasks, faces are most optimally found when presented in an upright orientation and a frontal pose, and when their natural colours and dimensions are preserved (Bindemann & Burton, 2009; Lewis & Edmonds, 2003; Pongakkasira & Bindemann, 2015).

However, the methodologies used to study face detection in previous research have varied greatly, in terms of the stimulus manipulations that have been applied to faces (e.g., inversion, scrambling, colour removal), the context in which faces are displayed (e.g., naturalistic scenes, visual search arrays), and the behavioural measures that have been used to measure face perception (e.g., eye-tracking, reaction times, search accuracy) (e.g., Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Hershler & Hochstein, 2005; Lewis & Edmonds, 2003). While these different methodologies have broadened the understanding of face detection, they also undermine the inferences that can be drawn across different studies, as it is difficult to separate the influence of different experimental approaches to study detection from the process itself. There is some initial evidence to suggest that the type of visual displays that are used to study detection not only influences the ability to find faces but also draws on different face processes (Bindeman & Lewis, 2013; Hershler, Golan, Bentin & Hoshstein, 2010). For example, whereas faces are found more slowly in profile than in frontal view in scene displays, performance is comparable for both when these are presented in blank displays, indicating that the mode of presentation affects the detection processes at work (Bindemann & Lewis, 2013). Furthermore, as these differences in finding frontal and profile faces arise in scene displays, there is a lack of clarity on which facial features drive these differences.

This thesis aims to better understand the process of face detection. The initial work focuses on whether people draw attention in naturalistic scene displays. Then, a series of experiments investigate how different contexts affect face detection. In these studies, information from the body and face are manipulated to define the optimal features for detection. I begin with a review of the relevant literature.

1.2 Attention to faces

One source of evidence for face detection comes from studies that have examined the allocation of *attention* to different stimuli. In this context, attention often refers to the preferential allocation of processing resources to one stimulus over another, which can lead to the faster or more accurate classification of a target or other information that subsequently appears in its location. Crucially, for attention to be allocated to a stimulus, it also has to be detected. Thus, studies of attention can provide insight into the visual conditions and facial features that might support detection. However, these attention processes are differentiated from detection in this literature review, as studies on the latter process typically involve specific instruction to *find* faces. Due to their nature, these processes interplay, and the distinction is not always clarified, and is used synonymously, e.g., being drawn to a face because it has been spotted versus spotting a face because of being drawn to it. In this section, studies of attention that are relevant to face detection are reviewed, focusing on the early development of face perception, and attention to faces in simple and increasingly complex displays.

1.2.1 Early Visual Face Perception

One source of evidence that faces compete strongly for attention comes from research with infants, which demonstrates an innate attentional bias towards faces. As infants cannot follow task instructions or exhibit social desirability bias that might lead to the prioritisation of social stimuli such as faces, they are impartial observers when presented with stimuli such as people or faces. Despite this, infants as young as 3-12 months shift their eye-gaze towards faces in visual arrays, where these are embedded among non-face objects, such as cars, shoes, animals and household items (Gliga, Elsabbagh, Andravizou & Johnson, 2009; Gluckmen & Johnson, 2013), and in more complex naturalistic scenes, where attention to faces can be indexed with eye movements (Kelly et al., 2019). Infants also view faces for longer than other objects (Gliga, Elsabbagh, Andravizou & Johnson, 2009; Gluckmen & Johnson, 2013). This indicates that these social stimuli can attract and hold attention from a young age, and signifies that faces hold importance in the early visual system of humans (Leppanen, 2016). This early attentional bias increases further with age (Di Giorgio, Turati, Altoè, & Simion, 2012; Kelly et al., 2019) and cannot be ascribed to stimulus saliency or "pop-out" effects (in which an item stands out due to a unique feature in the display), as it does not occur with inverted or scrambled faces, which retain the same visual information in a different structure (Brown, Huey & Findlay, 1997; Gluckmen & Johnson, 2013; Hershler & Hochstein, 2005). Therefore, the social importance of faces is either innate or quickly learned in early life and provides a powerful draw for attention.

1.2.2 Simple visual displays

In adults, similar attention biases have also been observed across a range of tasks that vary in complexity. The simplest versions of these tasks are dot-probe paradigms, in which observers are shown pairs of stimuli that are then subsequently replaced with a target probe in one of their locations. The speed and accuracy with which this probe is classified is then taken as an index of attention to the initial stimuli (see Figure 1.1). When faces are paired with non-face objects (such as a water-tap, a teapot, a wall clock, a train, a boat, and a dollhouse) in the initial display, the target probes are typically classified faster when these appear in the face locations (e.g., Bindemann & Burton, 2008; Bindemann et al., 2007). This effect occurs even with very short stimulus display times, which indicates that attention is drawn quickly to faces over other stimuli (Bindemann et al, 2007; Cooper & Langton, 2006). Alternatively, this attentional draw can be reversed, when faces negatively predict the location of the target (Bindemann et al., 2007). This suggested that attentional control can be strategically shifted away from face stimuli. However, attention was more readily shifted towards the face when cues were predictive, emphasising a face-specific attentional bias (Bindemann et al., 2007).

In addition, when the low-level physical attributes of face and non-face stimuli, such as luminance and contrast, are matched, this attentional bias disappears (Pereira, Birmingham & Ristic, 2019; Pereira, Birmingham & Ristic, 2020). This indicates that the mode of presentation also influences the allocation of attention to faces. Yet, once such attributes are adjusted, the unique features of faces that are important for the attentional draw might also be eliminated. Therefore, perhaps it is something about presenting naturalistic faces that draw attention, but variation in attributes (such as luminance and contrast) or task variation (how items are presented) can diminish such effects.

A noteworthy aspect of these paradigms is that some face characteristics do *not* appear to be important for these attention effects. For example, faces that are turned upside down (inverted) compete strongly with their upright counterparts in simple visual displays (Bindemann & Burton, 2008). Additionally, inverted faces can elicit a response bias when predictive of target cues. This suggests that some facial features, such as the internal arrangement of a pair of eyes above a nose and mouth, are not important for attention capture in these simple paradigms. This deviates from other tasks with faces, such as emotion perception (Savage & Lipp, 2015; Wegrzyn, Vogt, Kireclioglu, Schneider, & Kissler, 2017) and recognition (Ellis, Shepherd, & Davies, 1979; Jarudi & Sinha, 2003; Toseeb, Keeble, & Bryant, 2012), for which internal features are of central importance. However, an attentional bias is found with emotional expressions such as angry faces, suggesting that although internal feature configuration is not important for attentional capture in such tasks, they are still perceived and processed (Cooper & Langton, 2006).

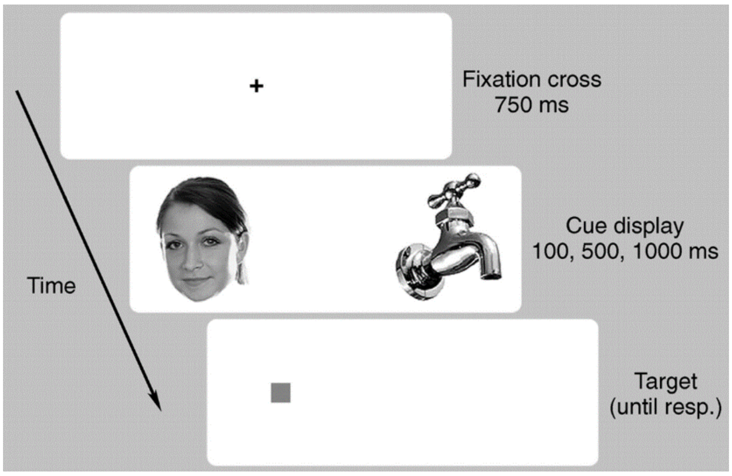


Figure 1.1. Example of the dot-probe task. Taken from Bindemann et al (2007).

1.2.3 Multi-item visual displays

A more complex paradigm that has been used to study the allocation of attention to faces is the change detection task or flicker paradigm. This method is also used to demonstrate the attention capture of faces, although it was originally employed to uncover how attention is allocated more generally in the visual field (Rensink, 2000; Rensink, O'Regan, & Clark, 1997). In this paradigm, two images flicker rapidly with one item or component of the image changing that observers must identify (usually appearing/disappearing, but can also be changing placement, colour etc.). A mask screen interjects these images so a direct comparison between the images cannot be made, rendering the task more difficult (Rensink, O’Regan, & Clark, 1997). Performance on this task indicates how attention is directed and represented within the visual display, as changes are found faster and with greater accuracy where attention is allocated (Rensink, 2000; Rensink, O’Regan, & Clark, 1997). In these paradigms, faces are presented in an array with 4-5 other objects, such as bananas, a toaster, a fan etc. (see Figure 1.2). The observers' task is to identify which item in the array is changing between the flickering images. Changes to faces are found faster than changes to other objects, indicating that attention is initially drawn to faces (Palermo & Rhodes, 2003; Ro, Russell, & Lavie, 2001). Furthermore, other items are slower to be found when faces are present (Crouzet, Kirchner & Thorpe, 2010; Langton, Law, Burton & Schweinberger, 2008; Little, Jenkins & Susilo, 2021). This suggests that these stimuli are particularly strong competitors for attention (but for exceptions, see Palermo & Rhoeds, 2003). However, in contrast to the simple visual displays in the dot-probe task, this face advantage appears sensitive to the visual characteristics of faces, as this effect is disrupted by stimulus inversion (Davies & Hoffman, 2002; Ro, Russell, & Lavie, 2001). This not only demonstrates that attention is predominately directed towards faces but also suggests that featural configuration might be important for such attentional draw in more complex displays.

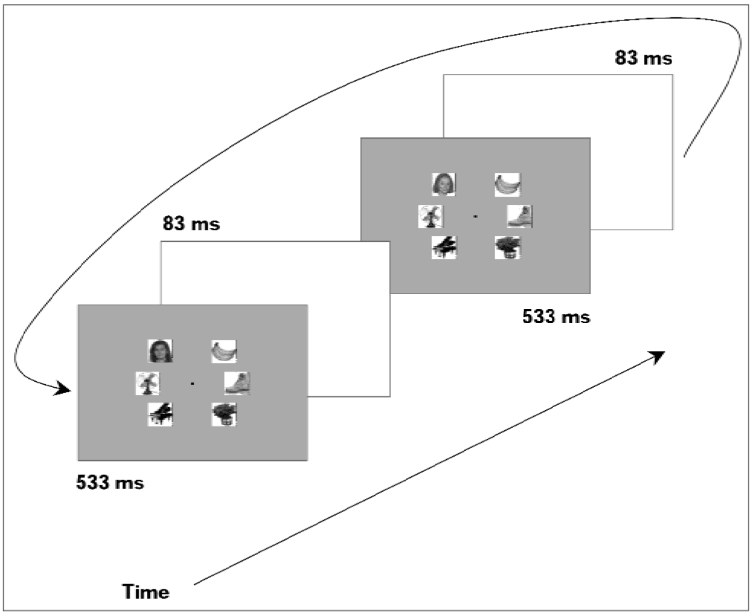


Figure 1.2. Example of the change detection flicker paradigm (from Ro, Russell, & Lavie, 2001)

However, due to the variations that can occur with array designs, research can present inconsistent findings depending on the variables used. For example, the face advantage disappears when more than one face is presented within the array, when the display comprises of less than four items, and when a single object is instead presented among faces (Palermo & Rhodes, 2003; Yang, Shih, Cheng & Yeh, 2009). This not only highlights how methodology plays a role in the conclusions that can be drawn from these studies but also notes that with such methods, faces may differentially exhibit "pop-out" effects depending on the surrounding visual *context*.

1.2.4 Complex visual displays

One type of visual context that is more representative of complex real-world circumstances than the simple displays of dot probe and flicker tasks are images of naturalistic scenes. Such displays are used in free viewing tasks, in which observers can naturally view the display, usually combined with eye-tracking as a measure of cognition and behaviour, and without instructions that purposefully direct attention towards specific stimuli of interest (see Figure 1.3). Different biases in looking behaviour occur during scene viewing (Tymkiw & Foulsham, 2020), one of which is to look at faces and person stimuli. For example, when scenes containing people are viewed, observers direct more fixations on the person regions in the scene than on any other areas (Birmingham, Bischof & Kingstone, 2008; Birmingham, Bischof & Kingstone, 2009). Moreover, the number of fixations on people increases as the number of people in a scene increases, demonstrating a draw towards these social stimuli (Birmingham, Bischof & Kingstone, 2008; Birmingham, Bischof & Kingstone, 2009). The use of saliency maps shows that saliency at the fixated person locations was low, and therefore did not account for the increased fixations. This indicates that the attentional draw towards people is due to their top-down, social importance (Birmingham, Bischof & Kingstone, 2009). Note that the person stimuli in the described research also covered a substantial proportion of the scene, which in itself has the potential to draw greater attention (see Figure 1.3) (Burton & Bindemann, 2009). However, when only a single person is present in complex scenes, observers locate these stimuli within just two fixations, indicating a fast attentional draw (Bindemann, Scheepers, Ferguson & Burton, 2010; Cerf, Frady, & Koch, 2009). Research also demonstrates that faces take priority over other objects that might engage attention strongly in everyday life, such as cell phones (Cerf, Frady, & Koch, 2009). Collectively, these studies converge with infant studies and attention experiments with more simplified displays, by indicating that faces are detected quickly and compete strongly for attention in natural occurrences.

Such free-viewing tasks also reveal the influence of task instructions, and how task variations affect looking behaviour during experiments (Birmingham, Bischof & Kingstonee, 2008). Greater fixations are directed towards the eyes when instructed to describe where attention is directed in a scene, compared to free viewing with no instructions (Birmingham, Bischof & Kingstonee, 2008). Additionally, fixations on person stimuli increase with the number of people within a scene, as well as in active scenes where people are depicted engaging with each other, demonstrating that both task demands and display play a role when evaluating such processes as attention. Although this emphasises the importance of utilising free viewing tasks to avoid the influence of task demands, eyes and faces were preferentially fixated, regardless of what is occurring in the scene (Bindemann, Scheepers & Burton, 2009; Birmingham, Bischof & Kingstonee, 2008). Eyes are even fixated when instructed not to (Laidlaw, Risko & Kingstonee, 2012). This preferential looking not only indicates the importance of faces in the visual system but also how internal features may contribute to this attentional bias.

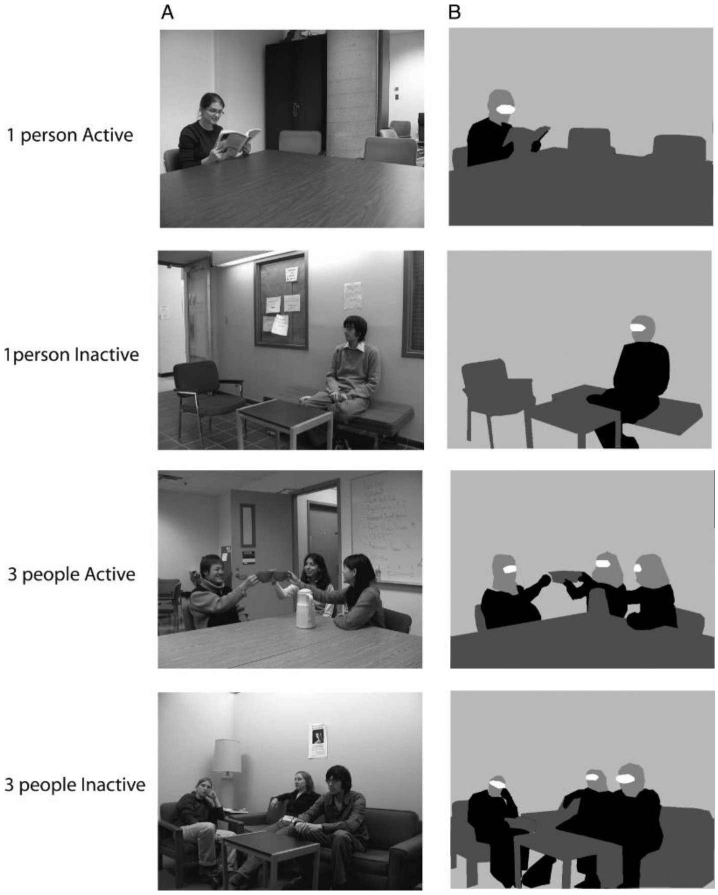


Figure 1.3. Example of presenting people in naturalistic scenes (from Birmingham, Bischof & Kingstonee, 2009)

Collectively, the presented evidence suggests that faces are strong competitors for attention, but the conditions under which this seems to occur are unclear. Variations in the display type (e.g., arrays), the task demands (e.g., relevant versus irrelevant to the task; the instructions given) and the presentation of faces (e.g., changes in the physical attributes) all influence whether this bias towards faces is found. Therefore, the research lacks a consensus on the drivers of this attentional draw, particularly in more naturalistic circumstances. The following section will focus directly on detection that provides more extensive evidence on the features that contribute to how faces are found.

1.3 Face Detection

The detection of faces is of primary interest in this thesis. Although the allocation of attention is one aspect of detection, here attention and detection are differentiated, as attention does not involve the explicit instruction to look for faces but allows inferences about detection to be made by virtue of the fact that faces normally have to be localised. However, detection focuses on the performance in localising a face within a display. The proceeding studies consider detection directly by investigating how quickly and accurately faces are found when observers are actively instructed to do so.

1.3.1 Rapid Face Detection

Faces seem to be detected rapidly compared to other objects in our environment. When instructed to search for specific items, faces are detected faster than other objects (e.g., vehicles) and animal faces (e.g., non-human primates) (Crouzet, Kirchner & Thorpe, 2010; Di Giorgio, Turati, Altoè, & Simion, 2012; Hershler & Hochstein, 2005; Maylott, Sansone, Jakobsen, & Simpson, 2021; Simpson, Maylott, Leonard, Lazo, & Jakobsen, 2019; Yang, Shih, Cheng, & Yeh, 2009). Furthermore, not only are faces found faster than other objects, but other objects are slower to be identified when faces are present (Langton, Law, Burton & Schweinberger, 2008; Ro, Friggel & Lavie, 2007), once again suggesting that faces distract attention (see Section 1.2). This effect predominantly occurs with upright faces, and not with inverted faces, indicating a processing bias that is not due to visual saliency, but that there is something important about upright, frontal view face stimuli.

This detection effect seems to be most effective when faces are presented upright, frontal and with intact features. Any variation on this template (e.g., inverting faces, changing face view or distorting internal features) seems to disrupt detection (Bindemann & Lewis, 2013; Hershler & Hochstein, 2005; Lewis & Edmonds, 2003). This initiated the debate as to whether internal features (e.g., eyes, mouth and nose) or external features (e.g., face shape, hair, skin colour) drive detection.

1.3.2. Internal Features

Distorting internal facial features (e.g., eyes, nose, mouth, eyebrows) is a key method in determining the important components that facilitate optimal face detection. These methods can change the face configuration without changing the perceptual saliency of the stimuli, therefore also giving insight into the "pop-out" versus social bias debate (Brown, Huey & Findlay, 1997). Such methods include scrambling, blurring, removing or obscuring features.

Scrambling faces, by rearranging the internal features, disrupts facial configuration to determine if internal features are necessary for face processing (see Figure 1.4). Detection performance decreases when features are scrambled compared to intact faces (Hershler & Hochstein, 2005; Kuehn & Jolicoeur, 1994). This effect was also found in both simple (blank displays) and more complex (visual arrays) displays. Evidence also comes from other forms of face processing, such as face categorisation, in which participants are asked to classify whether an item is a face, or an intact face. Scrambled faces are actively ignored and are not identified as faces when categorising whole faces (Donnelly, Humphreys, & Sawyer, 1993;

Taubert, Apthorp, Aagten-Murphy, & Alais, 2011). Therefore, the visual field distinguishes some facial details in order to identify face-like features. The extent to which faces are scrambled influences categorisation effects, too. Categorisation performance declines more with moderately scrambled (e.g., two features displaced) compared to highly scrambled (all features displaced) (Donnelly, Humphreys, & Sawyer, 1993). This is due to moderately scrambled faces still resembling full faces to an extent, making them harder to distinguish, yet highly scrambled faces can be differentiated easier due to the bigger difference in featural configuration. Distinguishing such features may also become more difficult during detection search tasks, due to the lack of visual acuity in the periphery (Brown, Huey, & Findlay, 1997; Kalloniatis & Luu, 2007; Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005; Westheimer, 1965). Therefore, face detection may also be sensitive to featural configurations in a similar way to categorisation, due to a lack of ability to identify features clearly in the periphery. Further evidence from EEG research shows that areas such as the fusiform gyrus have a greater N170 response (which is a neural response that occurs about 170 milliseconds after the presentation of a stimulus, and seems to be face-specific) to intact faces compared to scrambled, indicating features are important in face processing specificity (Cecchini et al., 2013). Therefore, the configuration of internal features seems to be an important component of a face detection template.

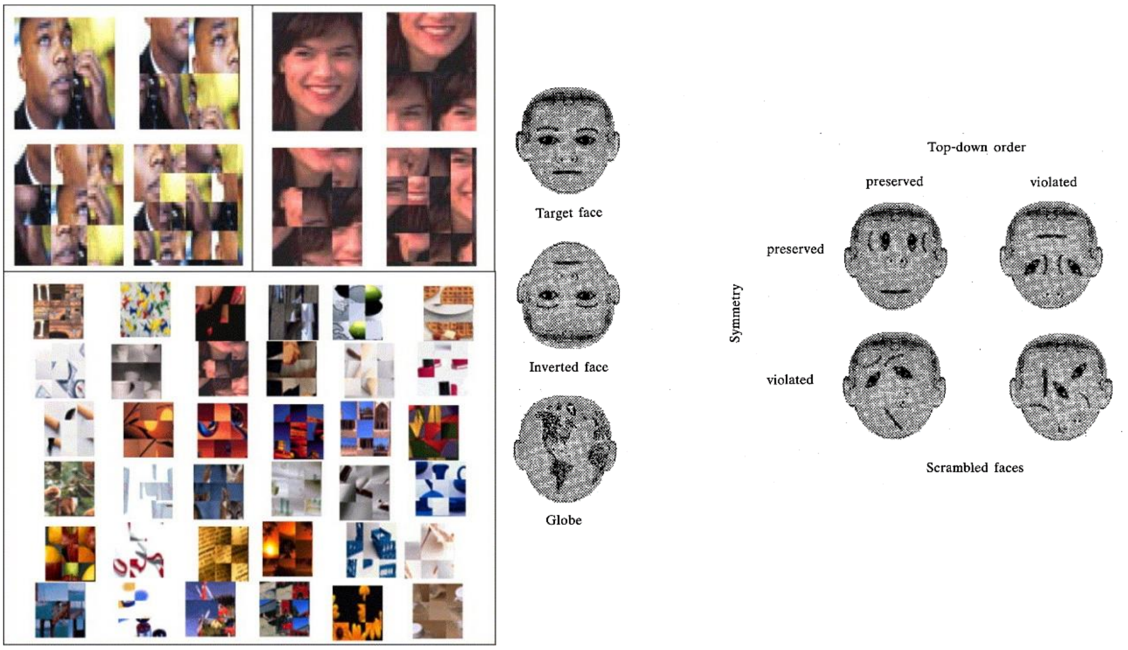


Figure 1.4: Examples of face scrambling (Left from Hershler & Hochstein, 2005; Right from Kuehn & Jolicoeur, 1994).

Obscuring, blocking, blurring or removing individual features (eyes, mouth, nose etc.) are other methods of face manipulation that affect face detection (see Figure 1.5). Blocking each individual feature separately causes a decline in detection performance, with the removal of information from the eyes eliciting the greatest detection cost (Lewis & Edmonds, 2003). This indicates that the eyes may be particularly important in detection, as well as other face processes such as recognition (Sadr, Jarudi, & Sinha, 2003). Furthermore, faces with all internal features blurred take longer to detect than intact faces suggesting that visible features are important in detection (Hershler & Hoshstein 2005; Lewis & Edmonds, 2003). Yet, blurred faces may also have the ability to elicit the N170 potential, suggesting they are still processed as faces (See section 1.4.1; Thierry et al., 2006).

Figure 1.5: Examples of blocking and removing features (Top from Lewis & Edmonds, 2003; Bottom from Sadr, Jarudi, & Sinha, 2003).

1.3.3 External Features

As an alternative to altering the configuration of features, more holistic, whole-face changes can be made to the face, such as colour, shape, and face view (Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Lewis & Edmonds, 2005; Pongakkasira & Bindemann, 2015). Adjusting some of these factors alters the saliency of faces (e.g., when changing colour and luminance), whereas others alter the visual structure (e.g., face rotation, inversion, shape), helping to differentiate the ‘pop-out’ effect from top-down processes.

Rotating faces (e.g., by 90º or 180º) keeps the visual saliency the same whilst also changing the general configuration. Inverting faces involves turning the face 180º, so it appears upside-down and is often compared to the detection of upright faces (see Figure 1.6). Inverting faces results in slower and less accurate detection rates (Brown, Huey, & Findlay, 1997; Lewis & Edmonds, 2003; Lewis & Edmonds, 2005), suggesting that something about the configuration of inverted faces disrupt detection. Though there is some debate over what is disrupting this process, i.e. whether it is the disruption of the external features or the re-arrangement of internal features (Brown, Huey, & Findlay, 1997; Lewis & Edmonds, 2003; Lewis & Edmonds, 2005). Additional evidence shows that inverted faces also take longer to break into visual awareness than upright faces during the phase unscrambling of images (Liu-Shuang, Ales, Rossion & Norcia, 2015). Although this type of detection does not involve search, it does represent the visual finding of faces.

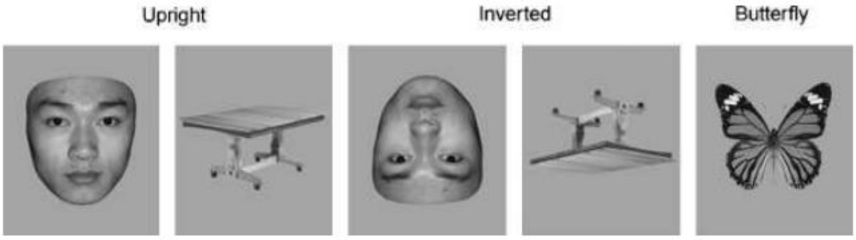


Figure 1.6. Example of upright and inverted face stimuli (from Gao et al., 2009)

EEG research gives further insight into the processing of inverted faces. Inverted faces produce greater N170 amplitudes than inverted face-like stimuli (such as phase scrambled face stimuli, car fronts that resemble faces, insects and

Arcimboldo paintings in which items are configured together to look like faces), demonstrating that inverted faces are still processed as faces, whereas this occurs less with other face stimuli (De Lissa et al., 2019; Nihei, Minami, & Nakauchi, 2018). This indicates that something about intact photo stimuli contributes to face processing, that more abstract faces cannot. Natural faces, therefore, hold specific content that may drive processes such as detection. The N170 increases and is delayed when faces are inverted (Gao et al., 2009). Additionally, inverted faces elicit decreased activity in brain areas associated with face detection (e.g., occipital-temporal regions) (Liu-Shuang, Ales, Rossion & Norcia, 2015), indicating that detection is disrupted by such stimuli. The overall indication is that upright faces capture important structural information that forms the predominant template for face processing, which alternative face presentations such as inverted faces do not contain.

Adjusting the colour, hue and luminance of faces also appears to play a significant role in face detection (Bindemann & Burton, 2009; Kemp, Pike, White and Musselman, 1996; Lewis & Edmonds, 2003; Lewis & Edmonds, 2005). Faces are detected faster and more accurately when in full colour compared to greyscale (Bindemann & Burton, 2009). This decline in performance when presented in greyscale occurs even when faces are only partially in greyscale. Performance declines further when faces are presented in unnatural skin tones, such as hue-inverted colours (Bindemann & Burton, 2009; Lewis & Edmonds, 2003). Colour, therefore, holds importance in face detection. Adjusting the hue and luminance of faces disrupts detection, with luminance seemingly having the most effect (Kemp, Pike, White & Musselman, 1996; Lewis & Edmonds, 2003; Lewis & Edmonds, 2005). As lighting and luminance adjust and shift in the real world, this could hold important applications in naturalistic detection. As the face structure is entirely intact, this implicates colour as a primary driver of detection.

Face detection is also influenced by the shape and size of the faces presented. Faces that are stretched either horizontally or vertically to form unnatural height-to-width ratios impair detection, implying that face shape is used as an identifier in detection (Pongakkasira & Bindemann, 2015). Here, detection performance also improves as face size increases (Bindemann and Burton, 2009). Size increase creates a more salient image, whilst also by default bringing the face closer to the central view, making it easier to identify. This is also confirmed by face eccentricity research, as although the N170 response seems to decrease at greater eccentricities, this effect is eliminated once size is accounted for – indicating low-level visual factors play a role in peripheral detection (Rousselet, Husk, Bennett & Sekuler, 2005).

Face view is a more naturalistic manipulation that does occur in daily experiences. Faces can be seen straight on frontal, in profile views, ¾ views, or have no visible features (see Figure 1.7). Frontal faces are detected faster and more accurately than profile faces (Burton & Bindemann, 2009). Three-quarter views are detected faster than profile views, but slower than frontal views (Bindemann & Lewis, 2013). This may also imply that colour and shape hold importance, as frontal faces consist of a standard oval, and is mostly skin tone, whereas variations on face view contain greater areas of hair which may also disrupt shape. However, there is some evidence that may suggest that the two eyes that are visible in frontal view may hold importance in detection. When faces are presented centrally on screen, eyes are fixated on preferentially for both frontal and ¾ view faces, but fixations land on the cheek region on profile faces, as the eyes are not visible (Bindemann, Scheepers & Burton, 2009). Therefore, the bias towards frontal faces was thought to be due to a propensity to fixate on the centre of a face, in which two eyes are in view (Bindemann, Scheepers & Burton, 2009). However, a pair of eyes may not be what is driving this bias, as detection is still faster for frontal faces even when only one side of the frontal face is presented with only one eye visible (Burton & Bindemann, 2009). Therefore there must be something that differentiates halved faces from profile or ¾ view faces. Another explanation poses that again, a colour shape template has involvement in the frontal face bias. Profile faces not only disrupt the face shape, but also the amount of skin tone visible, and therefore do not conform to this ideal template.

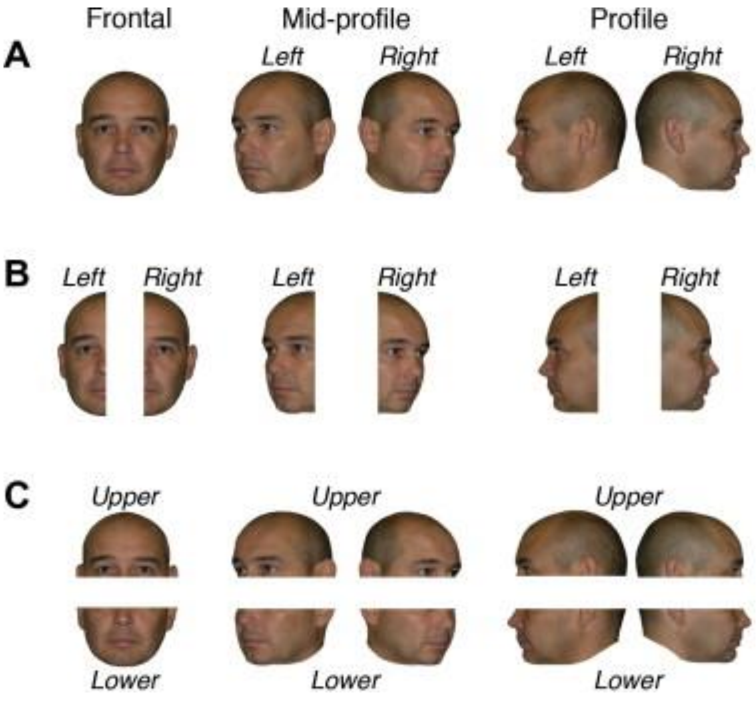


Figure 1.7. Example of frontal, three-quarter and profile face views in row A (from Burton & Bindemann., 2009).

1.3.4 Internal *versus* External Features

As both internal and external features have individually been shown to contribute to detection, there is a lack of consensus on which are essential. Although it may be suggested that they could contribute equally (e.g., Hershler & Hoshstein 2005), there is also a range of mixed evidence surrounding the topic. It is important to note the methods that are used may contribute to such uncertainty. For example, although research may be described as ‘detection’, the tasks may not involve the localisation of a face, such as using phase scrambled images (Liu-Shuang, Ales, Rossion & Norcia, 2015) or tasks that involve ‘searching’ blank backgrounds (Kuehn & Jolicoeur, 1994) (the influence of display is discussed further in section 1.4.2).

In some studies, detection performance was found to decline both when only internal or external features are separately blurred, suggesting that detection happens holistically and that internal and external features contribute equally (Hershler & Hoshstein 2005). However, it should be noted that external features are particularly difficult to blur, as the basic face shape still remains. This also does not consider other aspects, such as colour and luminance that have also been implicated (Lewis & Edmonds, 2003; Pongakkasira & Bindemann, 2015). It has been suggested that a colour-shape template, in which the basic shape and skin colour of frontal upright faces are maintained, seems to be a more significant driver of detection than internal features (Pongakkasira & Bindemann, 2015). This challenges opposing research that identify internal facial features as important components in face detection (Lewis & Edmonds, 2003). Yet this reasoning holds precedence as internal features are less salient in the periphery, making it hard to claim that they are identified in early detection processes (Brown, Huey, & Findlay, 1997; Kalloniatis & Luu, 2007; Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005; Westheimer, 1965).

Different types of face stimuli may also contribute to our understanding of this debate. Distinct face types range from real photographs, schematics, drawings, cartoons and pareidolia. Even within these categories, faces can be presented in different ways, such as photographs either presented in full or with cropped external features (e.g., hair, ears), or schematic faces being presented as Mooney faces (two-toned images). Real photographs of faces are one of the most accurate stimuli used to represent how we naturally process faces, yet even these lack the dynamics of in-person presence (Foulsham, et al, 2010; Itti, 2005). Possible implications can be considered with these different types, as all present different levels of detail and saliency which may be of importance in contributing to face detection (e.g., Sagiv & Bentin, 2001).

Schematic faces are depictions of faces using basic lines to represent facial features (see Figure 1.8). These faces highlight the key shapes that faces have (e.g. oval outline, 2 eyes, a nose, a mouth), whilst avoiding any recognition or identity of a person. Although such faces lack depth, they are still perceived as faces. Newborn infants show a preference for schematic faces with typical featural configuration compared to scrambled, demonstrating that they are categorised as face stimuli in early visual processing (Easterbrook, Kisilevsky, Hains & Muir, 1999; Johnson, Dziurawiec, Ellis & Morton, 1991).

Schematic faces depicting angry expressions are detected faster than other emotions, again, indicating that internal features may be salient enough to hold importance in detection (Dickins & Lipp, 2014; Lyyra, Hietanen, & Astikainen, 2014). Furthermore, these faces elicit the N170 component and produce stronger responses to emotions such as anger (Kreegipuu et al., 2013; Krombholz & Schaefer, Boucsein 2007; Lyyra, Hietanen, & Astikainen, 2014). This suggests that even faces that lack colour and depth are still processed as faces. Similarly to photographs of faces, removing features in schematic faces (such as the eyes, mouth, or outer circumference head shape) hinders detection (Calvo, & Nummenmaa, 2008; Sadr, Jarudi, & Sinha, 2003; Dickins & Lipp, 2014; Lewis & Edmonds, 2003). The use of schematic faces in face-processing research suggests that the basic lines that form the feature shapes are sufficient for detection. Yet this is not consistent with research addressing the importance of colour (Bindemann & Burton, 2009).

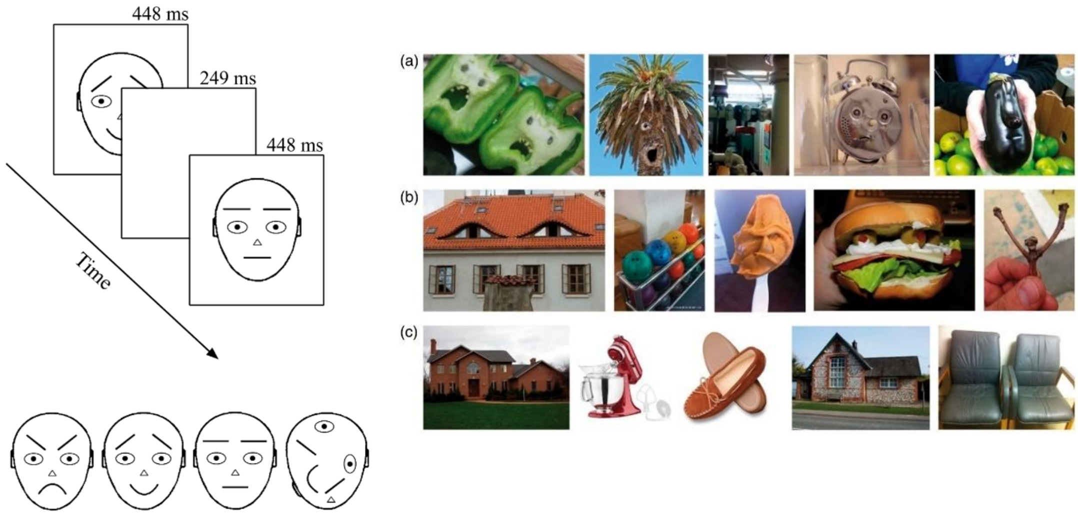


Figure 1.8. Example of schematic faces (Left. from Kreegipuu et al., 2013) and face pareidolia (Right. from Omer, Sapir, Hatuka, & Yovel, 2019)

Another interesting phenomenon that is used for researching face processing is face pareidolia (see Figure 1.8). This occurs when everyday objects are perceived as faces. Having items in the “eyes and mouth” configuration leads to objects being rated as more ‘face-like’ (Omer, Sapir, Hatuka, & Yovel, 2019). Objects perceived as faces have increased detection rates compared to non-face-like objects, but not as fast as real-photographed faces (Keys, Taubert, & Wardle, 2021; Takahashi & Watanabe, 2015). Brain activity, such as the N170 component and activation of the FFA, is also evoked by such faces (Hadjikhani, Kveraga, Naik, & Ahlfors, 2009; Wardle, Seymour, & Taubert, 2017; Zhou & Meng, 2020). This suggests that the basic configuration of features is enough to be detected and perceived as a face. Further to this, the belief that one might view faces in pure noise images creates this illusory effect that activates the FFA (Liu, et al., 2014). This indicates a top-down effect of face processing. Illusory faces are initially processed more like real faces than objects, with similar brain activation as viewing real faces, but are then transformed and recognised as objects (Wardle, Taubert, Teichmann, Baker, 2020). Therefore, the perception of such faces activates neural mechanisms consistent with face processing, suggesting that objects considered ‘face-like’ by the structure of the internal features alone are processed as faces. This indicates that it is such features that could drive detection (Alais, Xu, Wardle & Taubert, 2021). The implications for such effects are that a basic internal ‘feature-like’ template seems to contribute to face detection. The draw towards such basic ‘face-like’ stimuli implies external features and finer detail are less important.

Although all ‘face-like’ stimuli seem to be producing similar effects, such as all faces (photographs, painted portraits, sketches and schematic faces) eliciting the N170 component (Sagiv & Bentin, 2001) and infants being drawn towards both schematic and photographic intact faces (Macchi, Turati & Simion, 2004), there is evidence to suggest that these variations in presentation elicit some differences in processing (Keys, Taubert, & Wardle, 2021). For example, accuracy in identifying emotions is increased for ‘cartoonised’ faces, potentially due to more exaggerated features and less detailed complexity (Kendall, Raffaelli, Kingstonee, Todd, 2016). Similarly, gaze cueing exerts stronger effects with schematic faces compared to real faces (Hietanen & Leppanen, 2003). In both emotion and identity recognition tasks, more holistic processing occurs with real photographs of faces compared to schematic faces (Prazak & Burgund, 2014). These differences suggest that such stimuli may not directly equate to real-life face detection, so parallels in findings should be taken lightly.

Collectively, the type of faces used in face detection has implications for the conclusions that we draw. As all face types elicit face-specific responses, it directs us towards a ‘face-like’ configuration of eyes and mouth being responsible for face processing (Sagiv & Bentin, 2001). Though, it is of note that detection performance for photographic real faces has been found to be greater than other face-like types (Keys, Taubert, & Wardle, 2021; Takahashi & Watanabe, 2015), proposing that featural configuration cannot be the only driver of detection. As differing types of faces may be eliciting different processes, the presence of a colour-shape template may still hold importance in naturalistic detection. Converging with the previous discussion of the colour-shape template being an important driver in detection, it is harder to decipher the contribution of different face types to detection processes.

1.3.5 The Peripheral Advantage

Faces must first be detected in the periphery to facilitate this face advantage (Crouzet, Kirchner & Thorpe, 2010; Hershler, Golan, Bentin & Hoshstein, 2010). Faces, and even face-like stimuli, can be detected and categorised in peripheral vision whereas central vision biases occur with other items such as letters (Crouzet, Kirchner & Thorpe, 2010; Rieth et al., 2011). Faces are also categorised faster in the periphery than other objects (see Figure 9) (Crouzet, Kirchner & Thorpe, 2010; Hershler, Golan, Bentin & Hoshstein, 2010). Therefore, there is something specific about faces that can be identified in the periphery. Saccades towards faces occur even when they are task-irrelevant, such as searching for other objects (e.g., vehicles) (Crouzet, Kirchner & Thorpe, 2010; Little, Jenkins & Susilo, 2021). This also occurs when faces are inverted, when the image colour contrast is reversed, and when the amplitude and phase information of the image is adjusted (Awasthi, Friedman & Williams, 2011; Crouzet & Thorpe, 2011; Little, Jenkins & Susilo, 2021). This draw towards faces therefore appears to be face-specific but not due to orientation or colour (though note that effects are still stronger for ‘typical’ frontal faces). However, it is of note that such research does not provide a full picture of this face advantage. For example, although faces hold a bias when inverted, this is also when the other presented stimuli (e.g., cars) are also inverted. This lacks consideration of whether this face advantage holds against stimuli that are not inverted, which would provide a stronger case for face orientation not holding importance.

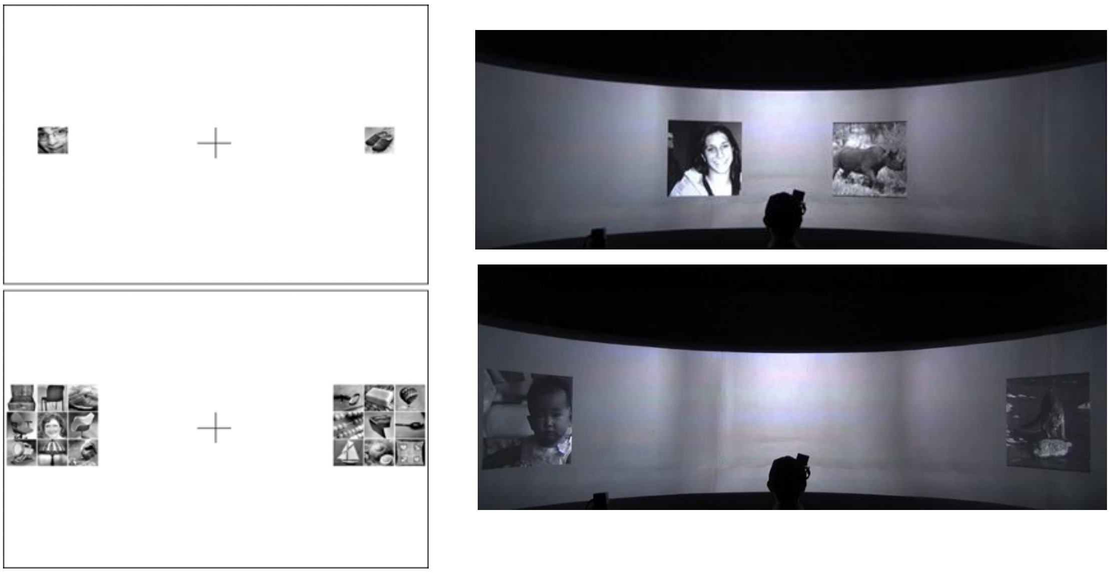


Figure 1.9. Examples of peripheral viewing paradigms. (Left from Hershler, Golan, Bentin & Hoshstein, 2010; Right: screen captures taken from a video from Boucart et al, 2016)

A peripheral detection advantage for faces is also evident in neural evidence. EEG research has revealed that the N170 component is elicited at the peripheral presentation of faces, rather than only centrally (De Lissa, McArthur, Hawelka, Palermo & Mahajan, 2019). Face-selective brain regions, such as the Fusiform Face Area (FFA), respond to faces presented in the periphery, albeit stronger when presented centrally (Levy et al., 2001). The perception of facial details is thought to elicit this stronger central response, yet this does not reflect peripheral face detection performance, in which detection is still faster than other items (Levy et al., 2001; Kanwisher, 2001). Therefore, despite a lack of visual acuity in the periphery, some facial detail must be visible in order to elicit such responses. The N170 component starts to diminish with greater eccentricities, however, this is thought to be attributed to cortical magnification, in which the faces cannot be ‘seen’ as clearly, rather than central bias. The low-level visual effect is overcome by increasing face size with eccentricity, returning the N170 to central-level responses (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005). As internal features are less visible at a greater distance in the periphery, this may provide evidence that a basic colour-shape template is sufficient to elicit a response. Presentation of faces at greater eccentricity in the periphery does not diminish the face detection effect. The face detection advantage over other objects is retained up to 80^o in the visual field during target search tasks and is sustained even in ‘crowded’ conditions in which images contain background information (see Figure 1.9) (Boucart et al, 2016). This demonstrates a detection advantage in our visual system for faces, whilst also highlighting that finer detail such as internal features may not drive these effects, as they are not necessarily distinguishable at such eccentricities.

The detection of faces in peripheral vision is prominent, yet the extent to which facial details are distinguished is still under question. Emotion detection can occur up to 40^o of eccentricity, however, gender does not yield the same advantage (Bayle, Schoendorff, Henaff & Krolak-Salmon, 2011). This suggests that some finer details may be visible in order to distinguish emotional characteristics, yet such detail is not enough to indicate gender (though this may be due to social constructs of gender rather than sex differences). Details such as gaze direction can be detected up to 6^o, however, a head orientation bias occurs at such distance, as at 3^o the details of the eye become harder to discern (Palanica & Itier, 2014). If such details can or cannot be distinguished in the periphery, then this gives insight into what may be observed during detection. The evidence points towards us being able to distinguish basic external facial features such as colour-shape in the periphery, but not necessarily finer details such as features. The features necessary for detection therefore may rely on external features more than internal details. However, there are methodological differences that may be exacerbating or hindering detection effects, contributing to the lack of consensus on the contribution of features. The next section will focus on how variations in methodology impact the conclusions that can be drawn about the cognitive processes underpinning face detection.

1.4 Methodological/Paradigm Variations

The wide variety of methods used have contributed substantially to our understanding of face detection. Certain methods can further our understanding of more naturalistic detection processes, such as the role of the body context. Differences in paradigms also provide different measures, for example, behavioural measures, brain activity or eye-tracking, to create a clearer picture of such processes. Yet the methods used have also created inconsistencies in findings and subsequently the conclusions that can be drawn. For example, the type of display or task may highlight differences in the importance of inverted faces, as noted when discussing attention (see section 1.2). These influences of methodological variations are explored further in the following sections.

1.4.1 The focus on faces

Most research on face detection, unsurprisingly, has focused on the faces themselves. Yet faces in the real world are not presented in isolation without other scenery and therefore focusing solely on the head neglects to consider the contexts in which we perceive faces. The whole body itself provides additional context when searching for people in our environment and therefore plays a role in person perception (Bindemann, Scheepers, Ferguson & Burton, 2010). Display context also provides perceptual cues and visual clutter that influence search and processing (Bindemann & Lewis, 2013). These must therefore be considered to build a representative model of face detection.

The body context demonstrates its importance in research using naturalistic scenes. A person is found faster in a scene when both the body and face are presented together, and although faces and bodies are detected at similar rates when presented separately, this is still slower than when presented together (see Figure 1.10) (Bindemann, Scheepers, Ferguson & Burton, 2010). First fixations show a body advantage, and the body receives the most fixations overall. However, faces show an advantage for second fixations and gain the most interest (Bindemann, Scheepers, Ferguson & Burton, 2010). Bodies can therefore be equally important in detection but faces ultimately hold attention once fixated. Body parts are also detected rapidly compared to objects (e.g., scissors, phones, plants etc.) and full bodies are detected faster than other stimuli, especially when presented together with a face (Bindemann, Scheepers, Ferguson & Burton, 2010; Downing, Bray, Rogers, & Childs, 2004). Body parts also distract attention in visual array tasks (Ro, Friggel, & Lavie, 2007). Therefore, the body likely plays a role in how people are detected in the real world, as they provide vital additional cues to the presence of a person.

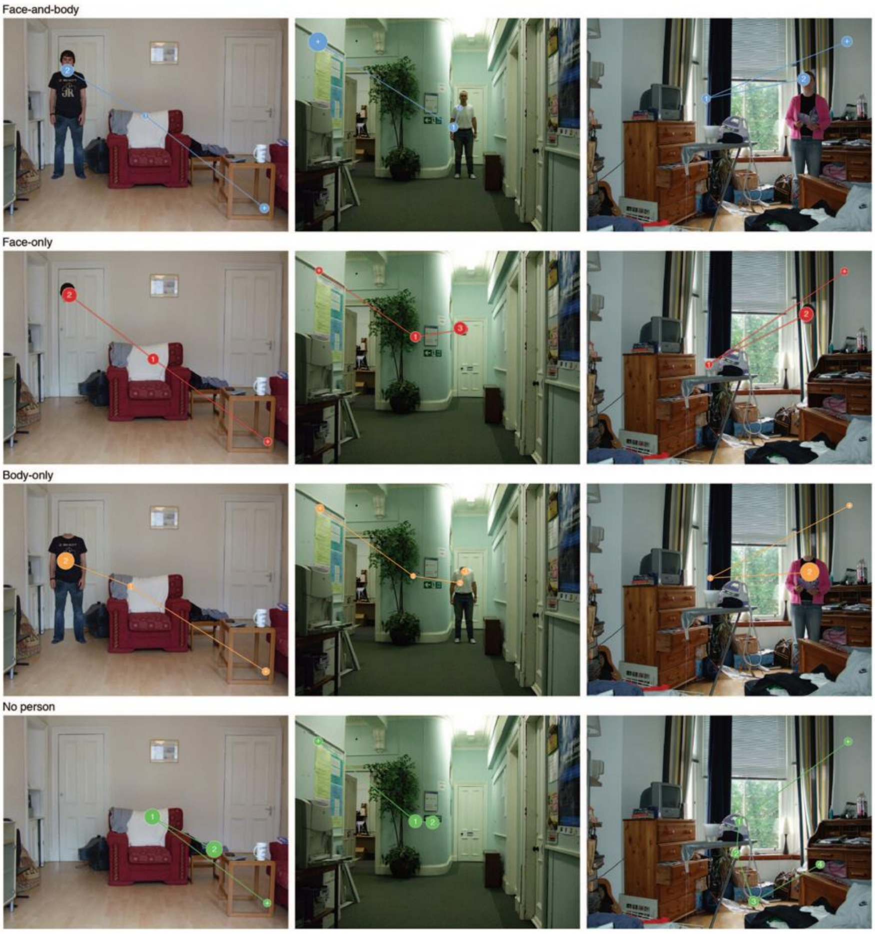


Figure 1.10. Examples of the full body, face only and body-only presentations in scene displays (from Bindemann, Scheepers, Ferguson & Burton, 2010)

As with faces, body areas also seem to have a neural basis, with specialised brain areas responding specifically to bodies that are distinct from face regions (Peelen & Downing, 2007). An area named the extrastriate body area (EBA; located in the focal region of the lateral occipitotemporal cortex) selectively responds to body parts and bodies but not to faces and objects (Downing, Jiang, Shuman, Kanwisher, 2001). The fusiform body area (FBA; located in the fusiform gyrus) also selectively responds to whole bodies and body parts (Peelen, Wiggett & Downing, 2006; Peelen & Downing, 2005; Schwarzlose, Baker & Kanwisher, 2005). The EBA is thought to be related to action processing, whereas the FBA overlaps with the FFA, so may have distinct category-selective neural representations (Downing, Peelen, Wiggett, & Tew, 2006; Peelen, Wiggett & Downing, 2006). The body therefore may need to be represented in different regions to process more body-specific volition (e.g., movement and motion). The role of specific body areas suggests that they also hold importance to our visual system, such as faces do, and therefore may contribute to the detection of people.

This is also shown in EEG research, in which bodies have similar responses to faces, but also may elicit a separate response. Hands can elicit the N170 component as well as faces (Kovacs et al., 2006). Bodies, particularly naked bodies, also elicit this component, whether the face is also present or not (Bernard, Content, Deltenre, & Colin, 2018; Hietanen & Nummenmaa, 2011; Stekelenburg & de Gelder, 2004). The N170 therefore may be considered a 'person-specific' component rather than just 'face-specific'. The human body also elicits an N190 component, differentiating it from the N170 face component (Thierry et al., 2006). Again, this may align with bodies and faces contributing separately to person detection. On the contrary, this difference in components may be due to how faces and bodies are presented. The aforementioned N170 research blurred faces so that they were not 'visible' (Thierry et al., 2006). However, the latter research removed heads entirely to find the N190 component. Blurred faces can therefore potentially trigger face-processing responses whereas bodies may incur a different response. Blurred heads eliciting such as response may hold support for a basic colour-shape template being dominant for face detection, as this template is enough to elicit face processing.

When the body context is considered, more explanation is provided as to how other manipulations contribute to or hinder face detection. Although face detection declines when scenes are scrambled (Lewis & Edmonds, 2003), the bodily information is also distorted which could contribute to the decreased performance, especially when considering that faces and bodies together contribute to detection greater than when presented alone (Bindemann, Scheepers, Ferguson & Burton, 2010). To further this, scrambling scenes means that perceptually and semantically, everything is out of place – affecting the processing of the display. Items are found faster in displays where they are located in feasible places (Davenport & Potter, 2004; Kaiser et al., 2019; Oliva & Torralba, 2007). Alternatively, items which are deemed semantically out of place may be drawn to faster due to standing out (Hollingworth & Henderson, 2000). A combination of bottom-up, contextual and top-down processing influences display processing and contributes to such effects (Torralba, Oliva & Castelhano, 2006). Therefore, the mode of presentation and the context of the display affect processing during detection tasks and therefore the conclusions that can be drawn from such research.

1.4.2 Methodologies and Display

A range of methodologies that are utilised in face detection research have already been discussed. Yet, there is little research on how these variations in methodologies influence conclusions on visual processing. The mode of presentation itself may have the potential to influence how faces are processed in visual displays, and therefore affect our understanding of such processes. Here, the focus is on the display type, predominately, blank displays, visual arrays and naturalistic scenes (Bindemann & Lewis, 2013; Hershler, Golan, Bentin & Hoshstein, 2010).

Firstly, blank displays allow for a simple presentation of stimuli, such as a single cropped face with no background (see Figure 1.11). This display is predominantly used in neuroscience methods such as EEG and fMRI to avoid noise and artefacts in the data (e.g., De Lissa, McArthur, Hawelka, Palermo, Mahajan, Degno, & Hutzler, 2019; Gao, Xu, Zhang, Zhao, Harel, & Bentin, 2009; Looser, Guntupalli, & Wheatley, 2013; Moulson, Balas, Nelson, & Sinha, 2011). The lack of interference from other stimuli means that this method is highly controlled in isolating the influence of the face itself. However, such displays do pose issues in research. Presenting faces in such a way removes the ‘search’ aspect of the task, influencing the process that is utilised, e.g., detection (finding a face) versus categorisation (classifying a face as a face). It is also unclear how such displays reflect real-world scenarios, as faces are not often presented in isolation without a visual context.

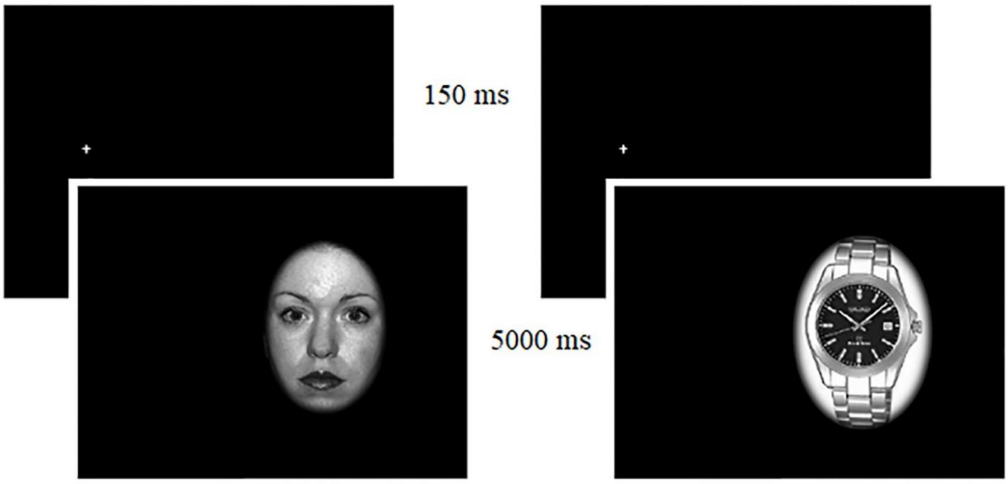


Figure 1.11. Example of cropped stimuli presented on blank backgrounds (from De Lissa, McArthur, Hawelka, Palermo, Mahajan, Degno, & Hutzler, 2019)

Visual Arrays are an alternative method, which has been mentioned previously in discussing other research such as the flicker paradigm, in which cropped faces are presented among a matrix of objects or other faces (see Figure 1.12). As each item is cropped, this allows them to be presented and processed individually without a background but can also be directly compared to the detection of other objects. This helps to determine which item is drawn to, or which is detected faster (e.g., Hershler & Hochstein, 2009; Langton, Law, Burton & Schweinberger, 2008; Meissner, Prüfer, Nordt, Semmelmann, & Weigelt, 2018; Ro, Friggel & Lavie, 2007; Simpson, Maylott, Leonard, Lazo, & Jakobsen, 2019). Search also becomes a part of such a task, therefore detection can be more clearly measured.

The presentation of arrays can vary greatly. Arrays can be structured either in a circular formation or grid form (see Figure 1.12). This also varies in the amount of background visible and how items are cropped. There is little research into how such variations affect search processes. The number of items in the array has influence, with a clearer face advantage emerging with an increased number of items (Hershler & Hochstein, 2005; Palermo & Rhoeds, 2003; Yang, Shih, Cheng & Yeh, 2009;). Additionally, the type of stimuli used has an influence. When an object is presented among an array of faces, as compared to a face among objects, the face advantage disappears (Palermo & Rhodes, 2003). The presentation of faces among objects may also produce an odd-one-out effect, in which faces are more interesting visually, such as being the only social stimuli in the display (Palermo & Rhodes, 2003). Furthermore, attention is drawn towards items which are deemed semantically out of place, such as a fire hydrant in a living room setting (Hollingworth & Henderson, 2000). Therefore, interest may drive the face advantage, rather than it being face-specific in the array context. Conclusions drawn from such experimental designs are therefore difficult to generalise to real-world scenarios (Palermo & Rhoeds, 2003; Yang, Shih, Cheng & Yeh, 2009).

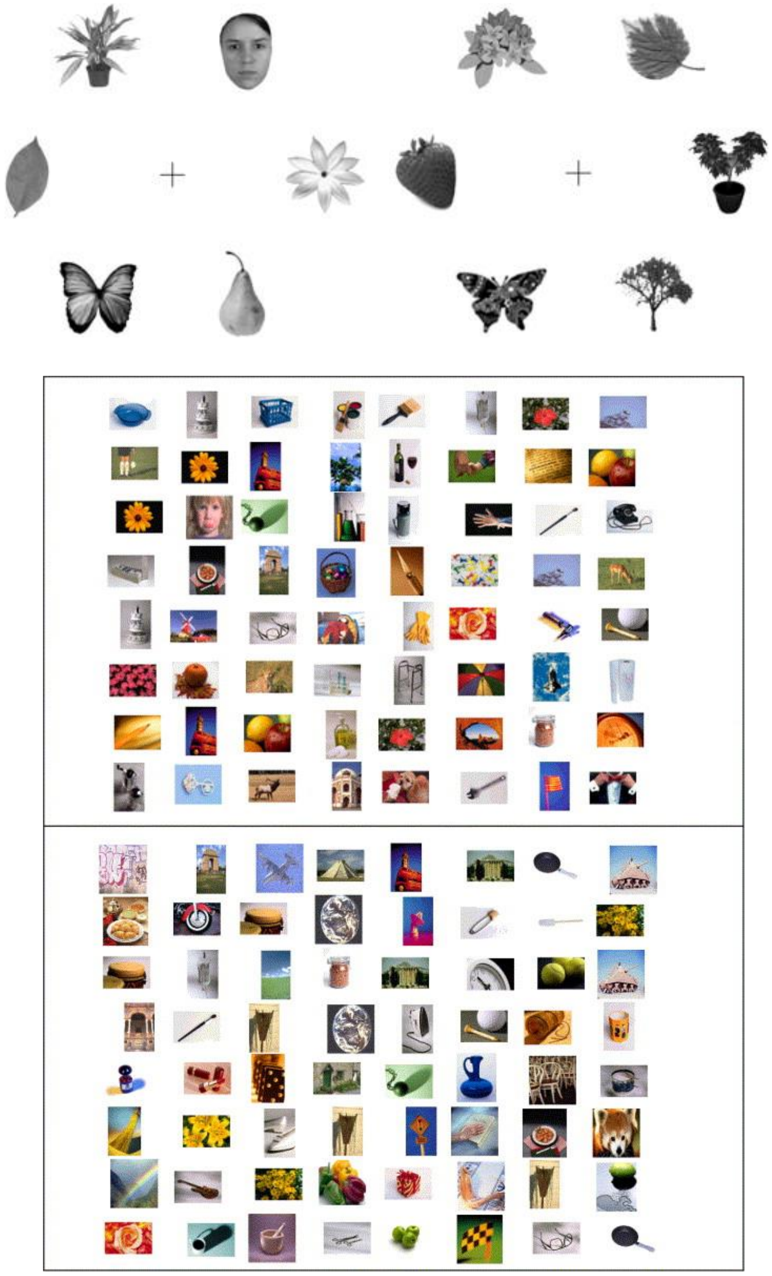


Figure 1.12. Example of a circular array of items (Top. From Langton, Law, Burton, & Schweinberger, 2008) and grid array of items (Bottom. from Hershler & Hochstein, 2005)

Finally, visual scenes (usually naturalistic photographs), are a better representation of how we view the real world and are therefore becoming more commonly used in research (see Figure 1.13). Embedding faces in such scenes creates a more naturalistic visual search that is potentially comparable to everyday visual processing (e.g., Bindemann et al., 2010; Bindemann & Burton, 2009; Burton & Bindemann, 2009; Kelly et al., 2019). Due to the nature of the stimuli, scenes vary in their complexity, in terms of colour, lighting, meaning, and context. Such variation influences scene processing, with different patterns of behavioural and neuronal responses depending on the level of complexity (Chai et al., 2010; Groen et al., 2018). The lack of standardisation across images makes it hard to directly compare stimuli, yet it does allow for more inference about real-world behaviour.

Such diversity in the display is likely to elicit differences in processing, and therefore, should be considered when forming inferences. Research has already started to demonstrate the impact of display types. For example, when presented at opposite ends of the periphery, faces and objects can be detected with equal efficiency if presented individually on blank backgrounds. However, a face advantage appears when these stimuli are presented within arrays (Hershler, Golan, Bentin & Hoshstein, 2010), indicating that detection functions differently depending on the surrounding stimuli. Furthermore, the detection of frontal and profile faces is comparable when presented on blank backgrounds, but a disparity appears when presented on scenes (Bindemann & Lewis, 2013). Therefore, display presentation can elicit different aspects of the detection process, or different processes altogether, that should be taken into consideration when drawing conclusions on face perception.

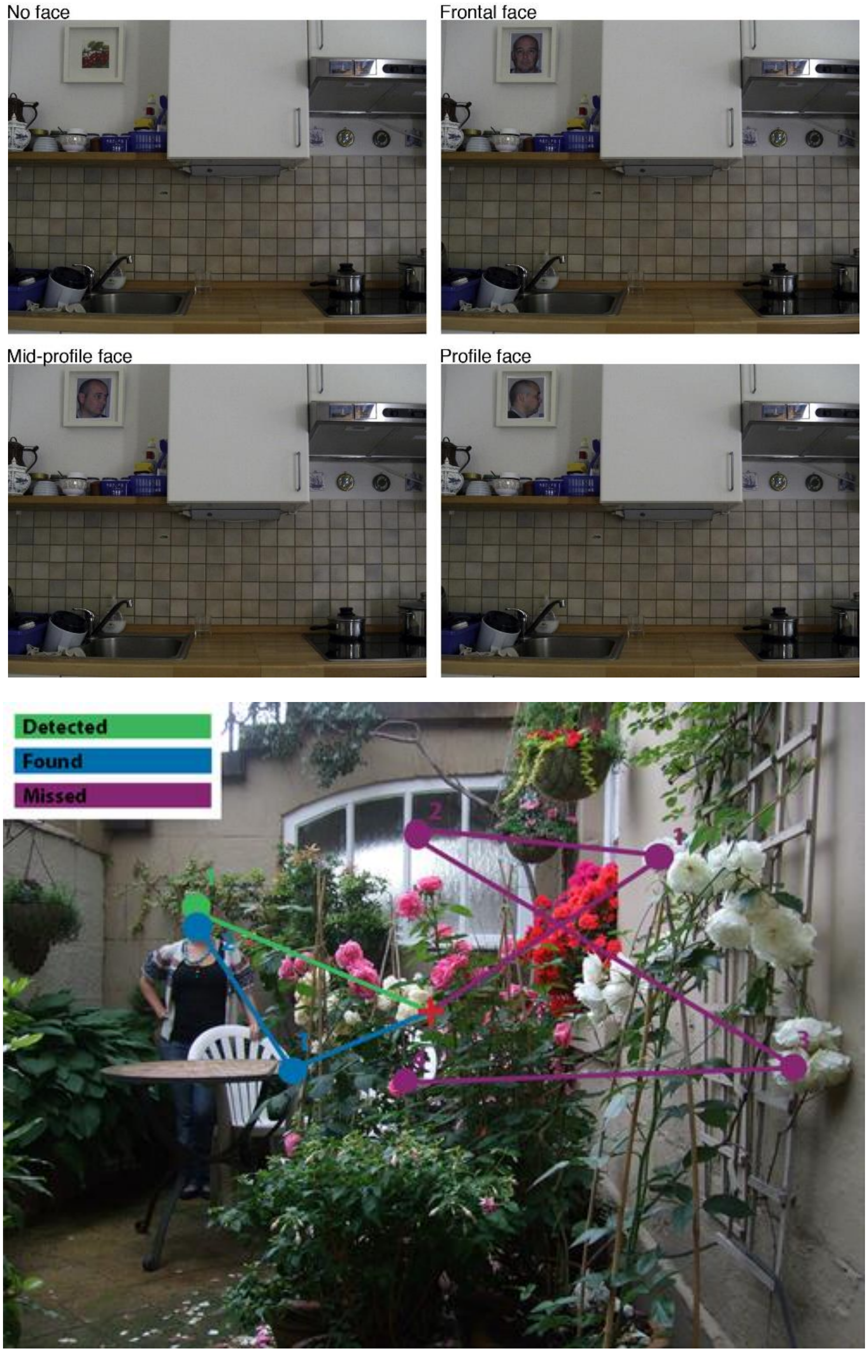


Figure 1.13. Examples of faces (Top. From Burton & Bindemann, 2009); and full-bodied people (Bottom. From Kelly et al., 2019) presented in naturalistic scenes.

Display context, therefore, influences the conclusions of detection research. The main difference is that complex displays require a search aspect to the task, to facilitate detection, which is not necessary for blank displays. Blank displays may be drawing on potential categorisation processes rather than detection, due to this lack of search (Bindemann & Lewis, 2013). The demands of the task can potentially influence the outcomes, for example, categorisation tasks have less effect on the cueing paradigms compared to detection tasks (McKay et al., 2021). This provides an explanation for the differences found for varying display types (Hershler, Golan, Bentin & Hoshstein, 2010), as face variations and objects are easier categorisations to make at initial presentation when not competing with other visual stimuli, yet when searched for, the noise from other items in a display, and cognitive templates may interfere with the search, making it easier to detect frontal faces over other items. This could explain how a basic colour-shape template may be prioritised in the visual system in search tasks (Pongakkasira & Bindemann, 2015) but is of less importance for other processes. In addition, this identifies how detection templates interact with task demands. Therefore, the use of scene displays that represent more natural visual perceptions of everyday scenes has a crucial role in the understanding of real-world processing.

Such differences in perceptions have important implications for research. As EEG and fMRI research present stimuli on simple displays to avoid noise in the data, they cannot sufficiently measure certain processes, such as detection, that involve aspects such as search. Such methods cannot, therefore, make precise inferences about processes of detection, but may be more representative of categorisation processes. It is also unclear from this as to the role of array paradigms. Due to the differences in array presentation, some may perform more closely to blank displays (e.g., few items in a circular array) compared to others closer to more complex scenes (e.g., grid array with many items). Such display differences should therefore be considered when forming conclusions from such paradigms. Another consideration is the use of behavioural measures in such research. A plethora of detection research relies on measures of reaction time and accuracy in order to form conclusions. These measures, however, involve more steps of processing (e.g. involving motor movement in order to make a key press) to initiate the action – whilst alternative measures such as fixations and saccades during eye-tracking have a more direct measure, furthering the understanding of the role of displays in searching for faces.

1.4.3 Eye-tracking

A range of measures are utilised to gain a further understanding of face processing. This ranges from behavioural measures such as reaction times and accuracy to eye-tracking methods such as saccades and fixations. Eye-tracking methods have their advantages, such as measuring earlier visual processes, rather than having to rely on physical responses (such as button pressing) that take longer to implement. Eye-tracking allows for more accurate identification of when a participant detects a person by looking at the time to first fixations, the length of fixations and initiation of saccades, allowing for more accurate inferences on the detection process.

Such methods have revealed faces are often fixated on within the first two fixations 65% of the time in naturalistic scenes (Cerf, Frady, & Koch, 2009). When the body context is also present in scenes, people were fixated within 2-3 fixations on average (Bindemann, Sheepers, Ferguson & Burton, 2010). Furthermore, when the body is present, it is more likely to be fixated before the face, and these fixations fall on the upper body. Thereafter, the face then receives more attention. This face bias has also been reinforced with research demonstrating preferential looking towards faces compared to text or cell phones, which are stimuli that are deemed societally important (Cerf, Frady, & Koch, 2009). Eye-tracking, therefore, provides valuable information on attention and detection performance in more naturalistic contexts.

Not only can eye-tracking be used as a measure, but also a manipulation to control how visual displays are viewed. Gaze contingency can be used so that parts of the visual display change or are masked as observers view the scene (see Figure 1.14). Foveal vision is tracked so that the area of fixation can be masked, or revealed as the observer is looking (Nuthmann & Canas-Bajo, 2022). Additionally, this allows for changes to occur as observers saccade towards particular areas of the image. Masking central vision allows for studying peripheral processing, as this method makes it impossible for the observer to fixate on any item in the image. Alternatively, the periphery can be masked to measure processing when only foveal vision is available (Van Diepan & d'Ydewalle, 2003; Van Belle, Graef, Verfaillie, Rossion & Lefevre, 2010; Van Belle, Lefevre & Rossion, 2015).

Currently, such methods have been used to demonstrate the importance of peripheral vision in search and attention. In visually complex scenes, inspection times increase when peripheral vision is masked, suggesting that scene processing has a reliance on peripheral viewing (Van Diepan & d'Ydewalle, 2003). The detection of non-face items (e.g. letters, objects) decreases when peripheral vision is unavailable, suggesting peripheral viewing holds importance over foveal vision for detection processes (Nuthmann & Canas-Bajo, 2022). Therefore, peripheral vision is vital in search tasks. Such methods have been applied to other aspects of face processing. During face recognition tasks, gaze contingency can distinguish between holistic and featural processing by masking peripheral vision so that only one feature at a time face be viewed, or masking central vision so that most of a face is visible apart from the area fixated (see Figure 1.14). This also highlights the inversion effect in which featural processing dominates compared to upright faces that are processed holistically (Richler & Gauthier, 2014). When gaze contingency is used to constrain peripheral vision, the inversion effect disappears as only specific features can be fixated (Van Belle, Graef, Verfaillie, Rossion & Lefevre, 2010; Van Belle, Lefevre & Rossion, 2015). This indicates that peripheral viewing is vital for holistic face processing. Emotion recognition of happy faces can occur in peripheral vision but are equally identified whether upright or inverted, suggesting emotion recognition relies on more featural processing (Calvo, Nummenmaa & Avero, 2010). Individual featural viewing hinders identity recognition abilities, with more fixations needing to be made and slower viewing (Maw & Pomplun, 2004; Van Belle, Graef, Verfaillie, Busigny & Rossion, 2010). This suggests that peripheral viewing is necessary for optimal processing during many face processes, and therefore may apply to detection.

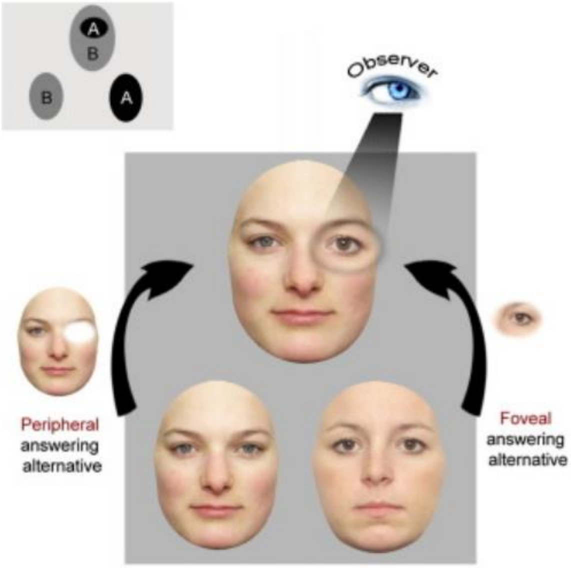


Figure 1.14. Example of the gaze contingency paradigm (From Van Belle, Lefevre & Rossion, 2015)

This collective evidence starts to form a picture of display and face processing, though this method has yet to be applied to face detection. Here, it is clear that peripheral vision is necessary for processing visual scenes and coupled with previously discussed evidence on face detection in the periphery, this suggests that peripheral vision is necessary for the detection of faces in complex displays. Additionally, this demonstrates how faces are processed holistically in the periphery, indicating that individual features in the periphery may not be necessary for detection processes.

1.5 Structure of this Thesis

This thesis explores the processes of face detection, whilst investigating the features necessary for optimal detection, and examines how methodology contributes to the conclusions that can be drawn from these studies. The first experimental chapter (Chapter 2) examines the attentional draw of person stimuli in naturalistic scenes, as well as how person frequency in tasks affects such attention. Across three experiments, a flicker paradigm is used to determine how person-presence draws attention away from detecting changes (Experiments 1, 2, & 3). Person frequency is also manipulated to determine the role of prevalence in attention (Experiment 3). If person-presence distracts attention spontaneously, then a decline in performance would be expected in person-present conditions, particularly when people appear more frequently in the task.

Chapter 3 furthers understanding of detection processes, by investigating the effect of display context on detection, whilst noting the facial features that drive this process. Four experiments present upright/frontal faces and manipulated (profile, rotated, inverted) faces on blank, array and scene displays to determine the differences in detection (Experiments 4, 5, 6 & 7). An additional experiment also uses Voronoi scenes to determine the effect of scene complexity (Experiment 8). If the detection of faces is comparable in blank displays, but disparities emerge between upright and manipulated faces in the array and scene displays, then this indicates that differences in processing faces arise depending on the display context. Detection is also expected to become increasingly difficult as display complexity increases, with a similar pattern arising in the disparity between detecting types of faces. Furthermore, the use of face manipulations can reveal the necessary features driving detection processes, such as external (face shape, colour, hair) or internal features (eyes, mouth, nose).

The final experimental chapter (Chapter 4) further investigates the role of internal and external facial features in peripheral detection. A gaze contingency eye-tracking method is used to manipulate the appearance and disappearance of internal features in peripheral vision (Experiments 9 & 10). Detection may be driven by a colour-shape template, in which case removing or adding features in the periphery should not impede detection performance. Additional experiments (Experiments, 11 & 12) aim to determine the extent to which features are visible in the periphery, by comparing whole and manipulated faces at different eccentricities. If performance differs between classifying featureless faces and when internal features are rotated, then this can determine whether internal features contribute to peripheral processing. This, therefore, starts to distinguish the features important in detecting faces in complex visual displays.

Chapter 2:

Spontaneous Attention to People in Scenes

Introduction

Person stimuli, such as faces and bodies, may draw rapid attention to facilitate the perception of important social information such as someone's identity and their attentional and emotional state. Faces and bodies engage attention faster than other objects (e.g., vehicles or butterflies) or animal faces (e.g., non-human primates) in visual search tasks and are also preferentially looked at, even when task-irrelevant, (e.g., when searching for other items) (Mayer, Vuong & Thornton, 2015; Simpson, Husband, Yee, Fullerton & Jakobsen, 2014; Langton, Law, Burton & Schweinberger, 2008; Crouzet, Kirchner & Thorpe, 2010; Little, Jenkins & Susilo, 2021; Cerf, Frady & Koch, 2009). Brain responses, such as activity in the fusiform face area (FFA), and the N170 component, which is a neural response that occurs about 170 milliseconds after the presentation of a stimulus, are also enhanced for faces and bodies compared to other visual stimuli. These effects persist when faces and bodies are presented in the peripheral visual field (De Lissa et al., 2019; Bernard, Content, Deltenre, & Colin, 2018; Hietanen & Nummenmaa, 2011; Stekelenburg & de Gelder, 2004). The sum of this evidence suggests that the human visual system is primed to be drawn towards people.

Such attentional gain can be dependent on task demands and stimulus design. For example, when the physical attributes of face stimuli, such as luminance and contrast, are equated with non-face stimuli (e.g., houses) in a dot-probe task, no attentional bias occurs (Pereira, Birmingham & Ristic, 2019; Pereira, Birmingham & Ristic, 2020). In addition, the mode of presentation of person stimuli also affects how conclusions can be generalised. Research often uses simplified displays, such as visual search arrays in which cropped faces are embedded among a small number of cropped non-face objects (e.g., Ro, Friggel & Lavie, 2007;

Langton, law, Burton & Schweinberger, 2008). In these studies, an attentional advantage for faces is typically found when presented among other objects. However, such stimuli lack the complexity of natural scenes, which can vary in colour, luminance, content and context (Bindemann & Burton, 2009; Burton & Bindemann, 2009; Kelly et al., 2019). In natural scenes, on the other hand, attention is also directed towards people (Bindemann, Scheepers, Ferguson, & Burton, 2010; Birmingham, Bischof & Kingstone, 2009). Yet differences arise in the information that draws attention in scenes compared to simpler visual displays, such as when faces are presented on otherwise blank displays or small scene regions. For example, whereas faces are found equally well in such simple displays irrespective of view, faces in a frontal view are detected more effectively in scenes than those in profile view (Bindemann & Lewis, 2013).

Almost all of these studies, however, have examined attention to people or faces in the visual field with paradigms that involve the explicit instruction to look for faces (e.g., Langton, Law, Burton & Schweinberger, 2008; Crouzet, Kirchner & Thorpe, 2010; Bindemann & Lewis, 2013) or have employed paradigms (e.g., dot probe) in which only a limited number of stimuli are displayed, so that the task-relevance of faces is clear (e.g., Pereira, Birmingham & Ristic, 2019; Pereira, Birmingham & Ristic, 2020). This raises the question of whether people draw attention in scenes *and* when participants are not explicitly looking for these stimuli. Some studies have tested this idea already, by comparing different task instructions, such as the active search for social stimuli and free viewing (Birmingham, Bischof & Kingstonee, 2009). This research indicates that people draw attention regardless of these task demands. However, the stimuli of this study comprised of indoor scenes, in which impoverished backgrounds were used and people occupied prominent scene regions.

Therefore, this research examines whether people draw attention when they are task-irrelevant and are presented in scenes that provide more complex viewing scenarios. To examine this, a Flicker paradigm was employed. In this paradigm, two images flicker rapidly, with one component of the image changing (usually appearing/disappearing), with a mask interjecting so that they cannot be directly compared (Rensink, O'Regan, & Clark, 1997). Performance on finding the change indicates how our attention is represented in visual displays (Rensink, 2000; Rensink, O'Regan, & Clark, 1997). This method has been employed to study attention to faces with simple visual displays, in which faces are shown alongside five non-face objects in a circular array. During a flicker, changes to faces are identified faster and more accurately than changes to other objects, indicating that attention is directed predominantly towards faces (Ro, Russell, & Lavie, 2001). This is also confirmed by a bias towards upright compared to inverted faces, as such effects are eliminated when faces are turned upside-down, which disrupts typical face processing (Ro, Russell, & Lavie, 2001; Wang, Miao, & Zhao, 2014; Yang, Shih, Cheng, & Yeh, 2009; Goodrich & Yonelinas, 2020; Davies & Hoffman, 2002). However, with simple visual displays, the method has also produced inconsistent results, whereby other studies have failed to replicate a face advantage (see Palermo & Rhodes, 2003).

The current research utilises the flicker paradigm with naturalistic scenes to investigate whether a bias towards person stimuli is found in these visually more complex settings. Naturalistic photographs containing people are presented, with a change occurring to an object somewhere in the image. In each task, observers were asked to indicate whether a change was occurring in the flicker images or not. The changes never occurred to the person present, to direct attention away from intentionally looking towards the person stimuli. If performance consistently declines when people are presented within a scene, then this gives insight into the attentional draw of faces and people.

Experiment 1

Experiment 1 examined whether people spontaneously draw viewers' attention when presented within naturalistic scenes. A flicker task was used to determine the influence of person presence when detecting changes within scenes. If people within a naturalistic scene draw attention spontaneously, then change-detection performance would decline in person-present scenes compared to person-absent scenes.

Methods

Participants

Forty-seven undergraduate students completed the experiment online (4 male, 43 female). A power analysis was conducted with G*Power based on a repeated-measures ANOVA (within-between interaction) with a medium effect size (f = .25, power = .95, number of groups = 1, and number of measurements = 4) and an alpha threshold of p = .05, leading to a suggested sample size of 36. This was adopted as the minimum sample size and advertised online for 3 weeks. The final sample represents all the sign-ups that occurred in this time period.

These participants ranged in age from 18-22 years, with a mean of 19.1 (*SD* = .89). Participants took part for course credit. All reported normal or corrected to normal vision. In all experiments, participants provided informed consent to take part.

Stimuli

The stimuli consisted of 40 photographs of naturalistic scenes that were taken around a university campus in daylight. All photos were taken outdoors and contained objects such as buildings, trees, stoplights, and cars. Twenty photos contained people naturally occurring within the scene so that they were different in face view, distance and the number of people. The remaining 20 photographs contained no people in view.

The images were sized to 1286 (w) x 965 (h) pixels at a resolution of 72ppi. Twenty of the photos (10 person-present, 10 person-absent) were edited using the graphics software Photoshop so that an item (such as signposts, footballs, trees, buildings etc.) was removed from the photo. For an example, see Figure 2.1. This allows for the image to flicker between the original image and the edited image so that a change was occurring. The remaining 20 photos were unchanged so that the same image flickered to create the no-change trials. The total of 40 trials consisted of 10 person-present change, 10 person-present no-change, 10 person-absent change and 10 person-absent no-change.

Procedure

Each trial began with a 1000 ms fixation cross before moving on to the stimulus display. Each trial cycled between the two image displays, with a blank white display separating them to create a mask. Both image displays would appear for 500 ms, with the blank display appearing for 300 ms following each image. This flicker would cycle until the participant responded to the task, by pressing the ‘c’ or ‘n’ keys on a computer keyboard to indicate ‘change’ or ‘no change’ respectively. Participants were instructed to indicate whether a change occurred in the image or not, and to respond as quickly and accurately as possible. After each trial, an interval screen was provided to avoid fatigue from concentration on flashing images. This remained until the participant responded to move on to the next trial by pressing the ‘space’ bar. Participants took part in all 40 trials, the order of which was randomised for each participant.

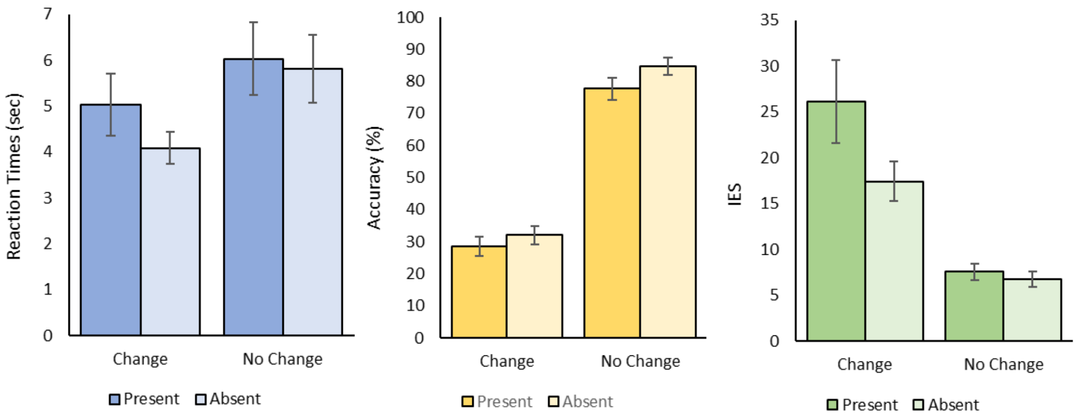
FIGURE 2.1. Example of person-present change stimuli with the original image (left) and the edited version (right).

Results

Thirteen participants were excluded from the analysis due to accuracy scores that fell more than two standard deviations below the sample mean (i.e. scoring below 41.77% accuracy) or reaction times that fell more than two standard deviations above the sample mean (i.e. responses above 12.85 s), leaving a final sample of 34. For the remaining participants, the mean correct response times and percentage accuracy were calculated. The average reaction times, accuracy and Inverse Efficiency Score (IES) data are illustrated in Figure 2.2.

FIGURE 2.2.

Reaction Times, Accuracy, and IES for Person-Present and Person-Absent Scenes in the Change and No Change conditions



Note. Average scores for reaction times, IES, and accuracy for person presence and person absence in the change and no change conditions. Error bars represent +/- 1 SE of the mean.

A 2(change, no change) x 2(person present, absent) repeated-measures ANOVA of these data was used to analyse the effects of change and person presence. For response times, ANOVA revealed that participants were faster to respond to change trials *(M* = 4.56 s, *SD* = 2.70) than no-change trials (*M* = 5.92 s, *SD* = 4.39), *F*(1,33) = 8.89, *p* = .005, η_p^2 = .21, as the search was completed once a change was found. This confirms that participants were adhering to task demands. However, ANOVA revealed no effect of person presence (*M* = 5.53 s, *SD* = 4.04) or absence (*M* = 4.95 s, *SD* = 2.91) on reaction times, *F*(1,33) = 3.19, *p* = .08, η_p^2 = .09, and no interaction between change and person presence, *F*(1, 33) = 1.54, *p* = .22, η_p^2 = .04.

For accuracy, ANOVA revealed that participants were less accurate on change trials (*M* = 30.3%, *SD* = 14.5) than on no-change trials (*M* = 81.2%, *SD* = 16.2), *F*(1,33) = 113.15, *p* < .001, η_p^2 = .77, due to being more likely to miss a change than falsely detect one. In addition, participants were more accurate on person-absent trials (*M* = 58.4%, *SD* = 7.56) than person-present trials (*M* = 53.1%, *SD* = 9.54), *F*(1,33) = 7.58, *p* = .01, η_p^2 = .19. No interaction was found between change and person presence for accuracy, *F*(1, 33) = 0.88, *p* = .36, η_p^2 = .03.

In a final step of the analysis, inverse efficiency scores (IES) were calculated to consider the trade-off between accuracy and reaction times (see Figure 2.2). Reaction times were divided by the proportion of correct responses. Higher IES indicate less efficiency in detection when the proportion of errors is taken into account. A repeated-measures ANOVA on these scores revealed less efficiency for the change condition (*M* = 21.78 s, *SD* = 16.30) than the no-change condition (*M* = 7.15 s, *SD* = 4.66), *F*(1,33) = 27.49, *p* <.001, η_p^2 = .45. Responses were also less efficient for person-present trials (*M* = 16.83 s, *SD* = 13.79) than person-absent trials (*M* = 12.09 s, *SD*= 6.79), *F*(1,33) = 4.66, *p* = .04, η_p^2 = .12. There was no interaction between change and person presence, *F*(1,33) = 3.15, *p* = .09, η_p^2 = .09

Discussion

Experiment 1 showed no initial effect on reaction times for detecting changes in the flicker task. This might reflect the time-dependent nature of the task, whereby flickers occur at specific times and participants would have to witness at least one loop to make a response. Therefore, due to the constraints of the flickers, it may be difficult to attain representative reaction time results. However, accuracy declined when a person was present within a scene. Additionally, when IES are considered, a difference in person-presence appears, which reflects the accuracy findings, with less efficiency for detecting scene changes in the presence of a person. Taken together, these findings suggest that people within naturalistic scenes spontaneously draw attention.

Experiment 2

Experiment 1 revealed an effect of attention capture by people in scenes, whereby this reduced the accuracy of change detection for non-face stimuli. However, accuracy levels were generally low and close to chance (of 50%), raising concern about the reliability of these effects. Experiment 2 therefore aimed to replicate the experiment with an easier version of the task. For this purpose, the images from Experiment 1 were cropped so that they were smaller, therefore making changes relatively larger and more salient to detect. If the findings of Experiment 1 are robust, then a decline in change detection performance should be observed when a person is present within these scenes compared to when people are absent.

Methods

Participants, Stimuli and Procedure

Forty participants completed the experiment (14 male, 26 female). These participants ranged in age from 18-39 years, with a mean of 25.8 (*SD* = 6.24). Participants were recruited from an online participation website (Prolific) and were paid a small fee to take part. All reported normal or corrected-to-normal vision.

The stimuli and procedure were the same as in the previous experiment except for the following changes. The same photographs from Experiment 1 were used, but one corner was cropped to make the image smaller, and therefore make the change appear larger and more visible, see Figure 2.3 for example. This is thought to improve the overall accuracy of the task as the change should appear more visible and require less search. How the image was cropped would vary based on where the change was occurring in the image. Photographs were then resized to match the stimulus dimensions of the previous experiment.

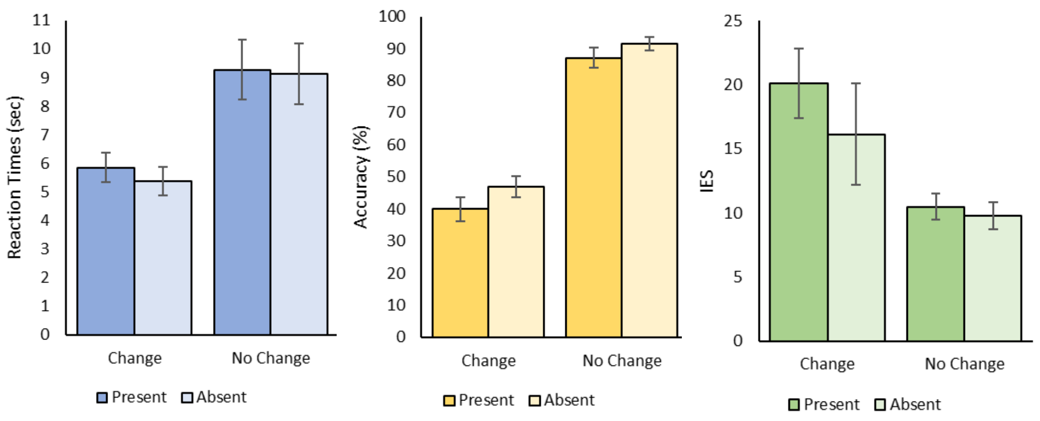
FIGURE 2.3. Example of the cropped stimuli, with the original images from Experiment 1 (top) and the cropped version for Experiment 2 (bottom).

Results

Five participants were excluded from the analysis due to accuracy scores that fell more than two standard deviations below the sample mean (i.e. scoring below 36.87% accuracy), leaving a final sample of thirty-five participants. For the remaining participants, the mean correct response times and percentage accuracy were calculated. The average reaction times, accuracy and IES data are illustrated in Figure 2.4.

FIGURE 2.4.

Reaction Times, Accuracy, and IES for Person-Present and Person-Absent Scenes in the Change and No Change conditions



Note. Average scores for reaction times, IES, and accuracy for person presence and person absence in the change and no change conditions. Error bars represent +/- 1 SE of the mean.

A 2(change, no-change) x 2(person present, absent) repeated-measures ANOVA was used to analyse the effects of change and person presence. For response time, ANOVA revealed faster reaction times for change trials *(M* = 5.62 s, *SD* = 2.52) than no-change trials, (*M* = 9.20 s, *SD* = 6.20), *F*(1,34) = 17.94, *p* < .001, η_p^2 = .35. This confirms that participants were adhering to task demands. However, there was no difference in response times between person-present (*M* = 7.57 ms, *SD* = 4.20) and person-absent trials (*M* = 7.26 ms, *SD* = 4.01), *F*(1, 34) = 1.21, *p* = .28, η_p^2 = .03 (see Figure 3), and no interaction between factors, *F*(1, 34) = .25, *p* = .62, η_p^2 = .01.

For accuracy, ANOVA revealed that participants were less accurate on change trials (*M* = 43.4%, *SD* = 18.9), than no-change trials (*M* = 89.3%, *SD* =

13.8), *F*(1, 34) = 151.90, p <.001, η_p^2 = .82. Additionally, responses were less accurate on person-present trials (*M* = 63.6%, *SD* = 15.1) than person-absent trials (*M* = 69.1%, *SD* = 12.3), *F*(1, 34) = 7.26, p = .01, η_p^2 = .18. No interaction was found between change and person presence, *F*(1, 34) = 0.48, p = .49, η_p^2 = .01.

Finally, IES was also calculated to consider the trade-off between accuracy and reaction times. A repeated measures ANOVA on these scores revealed less efficient responses in the change condition (*M* = 18.13, *SD* = 14.49) than in the no-change condition (*M* = 10.14, *SD* = 6.02), *F*(1, 34) = 7.86, p = .008, η_p^2 = .19. However, there was no difference in responses for person-present trials (*M* = 15.30, *SD* = 8.18) and person-absent trials (*M* = 12.98, *SD* = 11.62), *F*(1, 34) = .97, p = .33, η_p^2 = .03, and no interaction between change and person-presence, *F*(1, 34) = .47, p = .50, η_p^2 = .01.

Discussion

An Independent T-Test revealed that accuracy increased between Experiment 1 (*M* = 55.74%, *SD* = 6.53) and Experiment 2 (*M* = 66.36%, *SD* = 12.37), t(67) = 4.44, p < .001, d = 1.07. Therefore, Experiment 2 yielded similar results to Experiment 1, with the reduced difficulty of the task. Thus, change detection was less accurate in the presence of a person within the scenes than when no people were present, indicating that these stimuli were drawing attention away from the target regions. As in Experiment 1, this effect was not observed in response times, which is likely to reflect the lack of temporal sensitivity of change detection tasks. The same reasoning may also explain why an effect of person presence was not observed with inverse efficiency scores, which combine accuracy and response time measures.

Considering that a change never occurred to a person in these experiments, these findings suggest that people draw attention in scenes even when they are irrelevant to the target search. However, in Experiment 1 and 2, people appeared in 50% of trials. This frequent presentation of people potentially creates a perceived task relevancy, as higher target prevalence increases the likelihood that a target is identified, whereas low-prevalence targets are more likely to be missed by observers (Biggs, Adamo & Mitroff, 2014; Beanland, Le, & Bryne, 2016; Wolf et al., 2007; Goodwin et al., 2015; Menneer et al., 2010). Therefore, to provide a stronger test for the notion that people draw attention even when not task-relevant, the prevalence of person targets was manipulated in Experiment 3.

Experiments 3

Experiment 3 manipulates the frequency of person presence to investigate how task relevance and person prevalence interacts with attentional draw. Less frequent targets are more likely to be missed in search tasks (Biggs, Adamo & Mitroff, 2014; Beanland, Le, & Byrne, 2016; Wolf et al., 2007; Goodwin et al., 2015; Menneer et al., 2010). The relatively high-frequency rate of person presence in the previous experiments (50%) might therefore account for the consistent attentional draw (Biggs, Adamo & Mitroff, 2014). To examine this, three versions of the experiment were created in which a person was present 25%, 50% and 75% of the trials. If person prevalence interacts with their task relevance, then interference for detecting changes when a person is present should increase under high (75%) prevalence, and decrease under low (25%) prevalence.

To ensure that performance across these conditions was comparable, the stimulus set was also updated. In Experiments 1 and 2, the presentations of people within scenes varied unsystematically along several factors, such as the direction of a person's stance, face view, and their distance from the camera, which can influence person detection (see, e.g., Bindemann & Burton, 2009; Burton & Bindemann, 2009; Bindemann & Lewis, 2013; Or & Wilson, 2010). In Experiment 3, people were always presented facing forward and with a frontal face view.

Method

Participants

One Hundred Fifty-two undergraduate students completed one of three experiments in a between-subjects design (42 in the 25% prevalence condition, 61 in the 50% condition and 49 in the 75% condition). Variation in sample size occurred due to the number of sign-ups that were achieved once advertised online. One hundred twenty-four of the participants were female, twenty-four were male and 4 were non-binary, with an age range of 18-38 years (*M* = 19.55, *SD* = 2.88). Participants took part for course credit. All indicated normal or corrected-to-normal vision.

Stimuli and Procedure

One hundred sixty photographs of naturalistic scenes were taken around a university campus. In 80 of these photographs, one person would appear systematically facing forward in the scenes, to create a clear ‘frontal face’ person-present condition. The remaining 80 photos were person-absent. Images were sized to 1286 (w) x 965 (h) pixels at a resolution of 72ppi and were edited using the graphics software Photoshop so that an item (such as signposts, footballs, trees, buildings etc.) was removed from the photo. Each experiment contained 80 trials, with the amount of person-present trials varying. Example stimuli are displayed in Figure 2.5.

For the 25% condition, out of the 80 trials, only 20 contained a person within the scenes (10 trials in the change condition, 10 trials in the no-change condition), whereas the remaining 60 trials were person-absent (30 in the change condition, 30 in the no-change condition). The 50% condition contained an equal number of person-present and person-absent trials. Forty trials contained a person (20 change, 20 no-change), and the remaining 40 were absent (20 change, 20 no-change). For the 75% condition, 60 out of the 80 trials were person-present (30 change, 30 no-change), with the remaining 20 trials person-absent (10 change, 10 no-change). The trial procedure for this experiment remained the same as in the preceding experiments.

FIGURE 2.5. Example of person-present change stimuli for Experiment 3, with the original image (left) and the edited version (right).

Results

Thirteen participants (four in the 25% condition, six in the 50% condition and three in the 75% condition) were excluded from the analysis due to accuracy scores that fell more than two standard deviations below the sample mean (i.e. scoring below 43.31% accuracy in the 25% condition; 56.74% in the 50% condition; 40.88% in the 75% condition) or reaction times that fell more than two standard deviations above the sample mean (i.e. responses above 8.48 s in the 25% condition; 12.59 s in the 50% condition; 7.87 s in the 75% condition), leaving a final sample of one hundred thirty-nine. The mean correct response times and percentage accuracy were calculated for the remaining participants and are displayed in Figure 2.6.

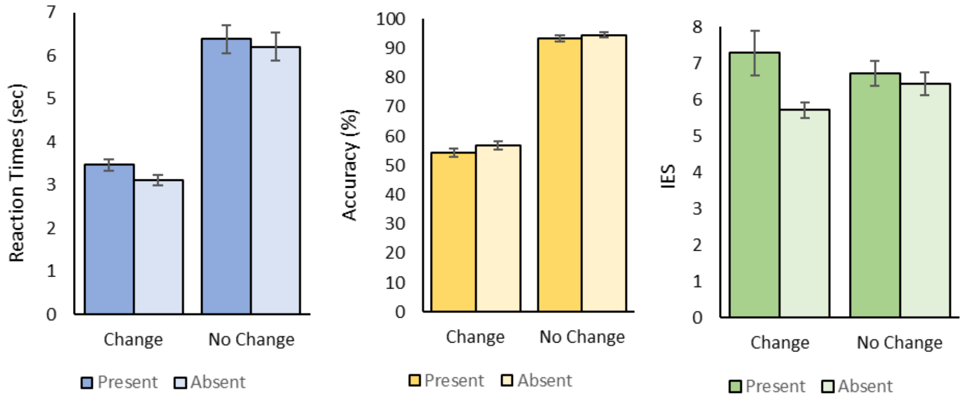
Pooled Analysis

In the first step of the analysis, the data from all three experiments was pooled to examine the overall effect of person presence on change detection. A 2(change, no-change) x 2(person present, absent) repeated-measures ANOVA was conducted for accuracy, response times and IES for the whole dataset. This data is represented in Figure 2.6.

For response times, ANOVA revealed that participants were faster to respond to change trials (M = 3.29 s, SD = 1.26) than no-change trials (M = 6.29 s, SD = 3.78), $F(1, 138) = 139.74$, $p < .001$, $\eta_p^2 = .50$, as the search was completed once a change was found. This confirms that participants were adhering to task demands. The ANOVA also revealed slower reaction times when a person was present in the scene (M = 4.92 s, SD = 2.47) compared to person absence (M = 4.65 s, SD = 2.37), $F(1, 138) = 13.34$, $p < .001$, $\eta_p^2 = .09$. No interaction between change and person presence was found, $F(1, 138) = 1.65$, $p = .20$, $\eta_p^2 = .01$.

FIGURE 2.6.

Reaction Times, Accuracy, and IES for Person-Present and Person-Absent Scenes in the Change and No Change conditions



Note. Average scores for reaction times, IES, and accuracy for person presence and person absence in the change and no change conditions. Error bars represent +/- 1 SE of the mean

For accuracy, ANOVA revealed that participants were less accurate on change trials (*M* = 55.7%, *SD* = 15.4) than on no-change trials (*M* = 93.9%, *SD* = 12.0), $F(1, 138) = 685.28$, $p < .001$, $\eta_p^2 = .83$, due to being more likely to miss a change than falsely detecting one. In addition, participants were more accurate on person-absent trials (*M* = 75.7%, *SD* = 11.52) than person-present trials (*M* = 73.7%, *SD* = 11.59), $F(1, 138) = 6.45$, $p = .01$, $\eta_p^2 = .05$. No interaction was found between change and person presence for accuracy, $F(1, 138) = 0.59$, $p = .44$, $\eta_p^2 = .004$.

In a final step of the analysis, inverse efficiency scores (IES) were calculated to consider the trade-off between accuracy and reaction times. A repeated-measures ANOVA on these scores revealed no difference in efficiency for the change condition (*M* = 6.51 s, *SD* = 4.15) and the no-change condition (*M* = 6.59 s, *SD* = 3.84), *F*(1, 138) = .03, *p* =.86, η_p^2 = .00. However, responses were less efficient for person-present trials (*M* = 7.01 s, *SD* = 4.08) than person-absent trials (*M* = 6.08 s, *SD*= 2.49), *F*(1, 138) = 10.09, *p* = .002, η_p^2 = .07. There was also an interaction between change and person presence, *F*(1, 138) = 4.65, *p* = .03, η_p^2 = .03.

Responses were less efficient for person-present change (*M* = 7.29, *SD* = 7.15) trials than person-absent (*M* = 5.72, *SD* = 2.57) change trials, *p* = .04, and person-present no-change trials, p < .001. However, there was no difference between change person-present trials and no-change person-present trials (*M* = 6.73, *SD* = 4.03), *p* = .86 and no change absent trials (*M* = 6.44, *SD* = 3.77), *p* = 60. In addition, efficiency for person-absent change trials was increased compared to person-present no-change trials, p = .04, but not for person-absent no-change trials, p = .17. Finally, responses were marginally less efficient when a person was present in the no-change condition, compared to in the no-change absent condition, *p* = .06.

Prevalence conditions

A series of 2(change, no-change) x 2(person present, absent) repeated-measures ANOVAs were conducted for accuracy and response times for each of the person prevalence conditions. The main effects and interactions for these ANOVAs are summarized in Table 2.1. The main findings can be summarised as follows. For response times, in all conditions, ANOVA revealed faster responses on change trials than no-change trials. Additionally, responses were slower in person-present than person-absent scenes under 25% person prevalence, but were comparable for person-present trials and person-absent trials at 50% and 75% person prevalence.

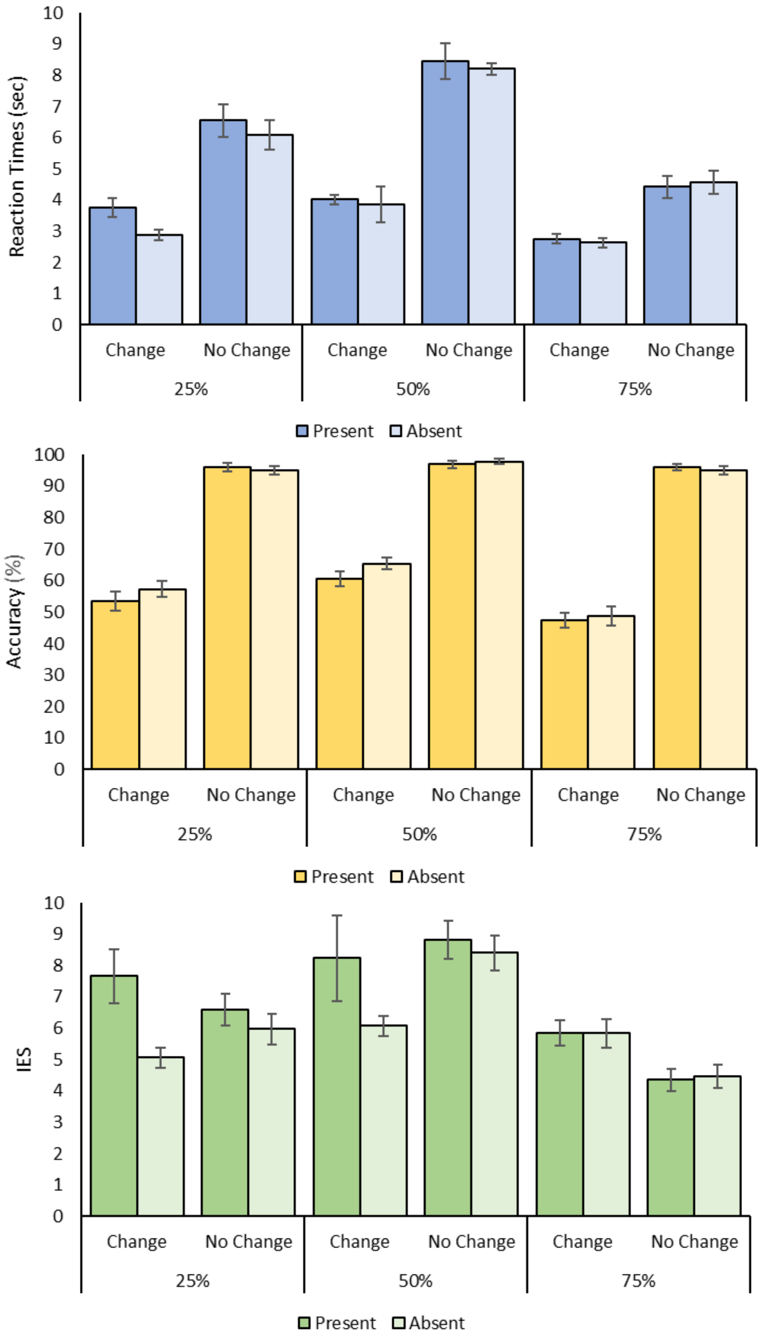
For accuracy, participants were less accurate on change trials than no-change trials in all frequency conditions, and participants were also less accurate on person-present than person-absent trials when person prevalence was at 50% or 25%. This effect was qualified by an interaction, whereby accuracy was lower for person-present change trials than person-absent change trials, *p* = .03, person-present no-change trials, *p* < .001, and person-absent no-change trials, *p* < .001. In addition, accuracy for person-absent change trials was also lower than for person-present no-change trials, *p* < .001, and person-absent no-change trials, *p* < .001. In contrast to the 50% prevalence condition, when people were presented in 25% and 75% of trials, accuracy for person-present and person-absent trials was comparable.

TABLE 2.1. A summary of comparisons for change and person presence, for RTs, Accuracy, and IES, for the 25%, 50% and 75% conditions in Experiment 3.

Comparison			25%	50%	75%
			F (1, 34) =	*F* (1, 54) =	*F* (1, 45) =
RTs	Change	No Change	68.06, *p* < .001***	80.73, *p* < .001***	40.96, *p* < .001***
	Present	Absent	16.47, *p* < .001***	3.47, *p* = .07	.03, *p* = .86
	Change *vs*	Presence	1.82, p = .19	.12, p = .73	2.24, p = .14
Accuracy	Change	No Change	213.05, *p* < .001***	238.61, *p* < .001***	263.32, *p* < .001***
	Present	Absent	1.00, *p* = .32	11.43, *p* = .001***	.33, *p* = .57
	Change *vs*	Presence	1.23, p = .28	4.50, p = .04*	.79, p = .38
IES	Change	No Change	.02, *p* = .90	2.04, *p* = .16	11.35, *p* = .002**
	Present	Absent	20.28, *p < .001* ***	3.86, *p* = .06	0.03, *p* = .85
	Change *vs*	Presence	1.23, p = .28	1.73, p = .19	.08, p = .78

FIGURE 2.7.

Reaction Times, IES, and Accuracy for Person-Present and Person-Absent Scenes in the Change and No Change conditions



Note. *Average scores for reaction times, IES, and accuracy for person presence and person absence in the change and no change conditions. This is represented across the 25%, 50% and 75% conditions. Error bars represent +/- 1 SE from the mean.*

Finally, analysis of inverse efficiency scores revealed that responses were less efficient for the change condition than the no-change condition under 75% prevalence. In addition, responses were less efficient for person-present than for person-absent trials under 25% prevalence. No other effects reached significance.

Discussion

Experiment 3 replicates the main finding of Experiments 1 and 2, by showing that change detection in scenes was less accurate when people were present compared to when these were absent. As in previous experiments, this effect was observed when people were presented in 50% of trials. Experiment 3 extends these findings to a new stimulus set, in which faces and bodies were always presented in a frontal view, suggesting that this is a robust effect.

In contrast to the preceding experiments, the prevalence of people in scenes was also manipulated, to examine whether people draw attention spontaneously in these scenes despite their task irrelevance. This was based on the notion that, although a change never occurred in a person in the scenes, the relatively high-frequency rate of person presence in the previous experiments (50%) might have served to draw attention to these stimuli in itself (Biggs, Adamo & Mitroff, 2014). Contrary to these predictions, the result showed that change detection was impaired under low person prevalence (25%), but not under high prevalence (75%). In contrast to the 50% prevalence condition, this effect was evident in response times and inverse efficiency scores. This finding seemingly contradicts previous findings on object target frequency (Biggs, Adamo & Mitroff, 2014; Beanland, Le, & Bryne, 2016; Wolf et al., 2007; Goodwin et al., 2015; Menneer et al., 2010), whereby people should garner greater attention due to being present more frequently throughout the task. An alternative explanation for this finding could be that under high person frequency, it becomes apparent that changes do not occur to the people in scenes, leading observers to reduce the allocation of attention to these particular regions (see Bindemann et al., 2007).

General Discussion

The current research shows that people draw attention spontaneously in naturalistic scenes. When a person is present in a real-world display, change detection performance in a flicker-task declines for both accuracy (Experiments 1, 2, 3) and response times (Experiment 3) compared to when a person is absent from a scene. This indicates that observers' attention is drawn towards people in scenes, causing a greater likelihood of change-detection errors and longer search times. These results are consistent with previous flicker research that demonstrates an attentional bias to faces (Ro, Russell, & Lavie, 2001) and extends these findings to visual search in natural scene displays.

The current experiments further these results, by demonstrating that the attentional engagement of people depends on the frequency with which these stimuli appear in scenes. In Experiment 3, change detection was slowed down by person presence when people appeared in 25% of scenes, but not with 75% prevalence. This diverges from visual search with non-face targets, in which less frequent targets are more likely to be missed by observers (Biggs, Adamo & Mitroff, 2014; Beanland, Le, & Bryne, 2016; Wolf et al., 2007; Goodwin et al., 2015; Menneer et al., 2010). In the current study, increasing person prevalence may have served to emphasise the irrelevance of these stimuli in detecting the changes in scenes, leading observers to actively shift attention away from these regions (see Bindemann et al., 2007). It is of note, however, that the person within the scene was not the participant's target for the task, but rather the change was the target. Yet the research is still in line with previous research on attentional draw when search is purposefully directed towards faces or person stimuli (e.g. Ro, Russell & Lavie, 2001).

Findings in attentional research had often been inconsistent due to variations in the nature of the task, such as odd one-out effects, cognitive load, semantic placements and display types (Hollingworth & Henderson, 2000; Palermo & Rhodes, 2003; Bindeman & Lewis, 2013). As naturalistic scenes were used rather than visual search arrays, the cognitive load effect is closer to how visual scenes are represented in real-world scenarios and therefore is less of a concern in these experiments. The use of natural photographs additionally removes any semantic 'out of place' effects (Hollingworth & Henderson, 2000; Palermo & Rhodes, 2003), furthering the evidence towards attention being a spontaneous process. Such scenes further distinguish this research from previous works. Complex displays elicit greater search effort and are more relevant to real-world experience (e.g., Bindemann, Scheepers, Ferguson, & Burton, 2010; Bindemann & Burton, 2009; Burton & Bindemann, 2009; Kelly et al., 2019). More representative conclusions can therefore be drawn on the allocation of attention. Although scene displays also exhibit greater variability, creating more confounding variables, two different stimulus sets were used across the experiments reported here, increasing confidence that the reported results are robust.

Investigating the role of full-bodied people, rather than faces in isolation, also differentiates this research from other studies. The role of the whole body is not often considered in attentional bias, yet we do not see faces in isolation in the real world. The current research demonstrates attentional draw occurs with whole bodies, which aligns with other research that emphasises the importance of the body (Bindemann, Scheepers, Ferguson, & Burton, 2010; Fletcher-Watson, Findlay, Leekam, & Benson, 2008). Again, this also speaks to the importance of task-dependent context. When faces are presented in arrays or are placed in scenes, they may be considered out of place, either visually 'popping out' or being semantically different from other items (Palermo & Rhodes, 2003; Hollingworth & Henderson, 2000). By presenting a full person in natural photographs, the current experiments remove these task-dependent factors.

In summary, the current experiments suggest that people draw attention spontaneously in natural visual scenes, as evidenced by impaired change detection. However, these effects persist only when people occur on half or less than half of all trials. When people are appearing often but are found to be task-irrelevant, they may be dismissed in favour of the task's focus. This gives insight into how attention is directed within scenes, indicating that person stimuli draw attention spontaneously but not automatically in complex visual displays.

Chapter 3:

Understanding Face Detection in Visual Arrays and Real-World Scenes

Introduction

People have a clear propensity to draw spontaneous attention in visual scenes, as observed in Chapter 2. When presented in naturalistic scenes, the presence of a person is able to distract observers away from finding changes in a flicker task. This was consistent with previous research, in that performance on finding changes increased when they occurred to faces compared to other objects, indicating attention is initially directed towards the person stimuli (e.g. Ro, Russell & Lavie, 2001). The previous chapter built on this explanation, by ensuring that the person present was irrelevant to the tasks, and therefore directing attention towards such stimuli would not have benefit to finding the change. This revealed the spontaneous nature of the attentional draw. Additionally, the use of naturalistic scene displays allowed for more representative conclusions to be drawn on the allocation of attention, as well as controlling for factors such as cognitive load or semantic 'out-of-place' effects, that have been shown to influence previous tasks (Hollingworth & Henderson, 2000; Palermo & Rhodes, 2003).

The previous Chapter also extended this further by considering the role of person frequency. Previous research has suggested that less frequent targets are more likely to be missed by observers (Biggs, Adamo & Mitroff, 2014; Beanland, Le, & Bryne, 2016; Wolf et al., 2007; Goodwin et al., 2015; Menneer et al., 2010). Chapter 2 examined this by manipulating the person frequency, by presenting people in 25%, 50% and 75% of trials. This revealed that identifying changes in scenes declined when people were presented infrequently (25% and 50%), yet person presence did not influence detecting changes when presented in 75% of trials. This indicates that increased person prevalence emphasises the irrelevance of the person within the task, and therefore attention can be actively shifted away from these regions.

The impaired detection of changes with person stimuli is distinguished from previous research with the use of full-bodied people. This allowed for a more naturalistic depiction of stimuli, whilst also acknowledging that the body may play a role in such attentional processes (e.g., Bindemann, Scheepers, Ferguson, & Burton, 2010). The use of such stimuli demonstrated that full-bodied people, depicted in naturalistic scenes can draw attention spontaneously. Yet, this does not provide an explanation as to what is specifically driving this person bias, and why people are detected easily in the visual field.

The face provides a primary candidate for such person detection. Faces are located rapidly in the visual field (Crouzet, Kirchner, & Thorpe, 2010; Crouzet & Thorpe, 2011) and are detected faster than animal faces and non-face objects (e.g., Di Giorgio, Turati, Altoè, & Simion, 2012; Maylott, Sansone, Jakobsen, & Simpson, 2021; Simpson, Maylott, Leonard, Lazo, & Jakobsen, 2019; Yang, Shih, Cheng, & Yeh, 2009). Despite providing much smaller visual cues, faces are also detected as quickly as human bodies, and are fixated preferentially, indicating a prominent role in person perception (Bindemann, Scheepers, Ferguson, & Burton, 2010). This efficient detection supports other tasks with faces, too. For example, faces in the visual periphery compete strongly for cognitive resources, such as those necessary for person identification, even when they are task-irrelevant (Bindemann, Burton, & Jenkins, 2005; Cerf, Frady, & Koch, 2009; Jenkins, Lavie, & Driver, 2003; Langton, Law, Burton, & Schweinberger, 2008).

Some of the visual features that support face detection have now been identified. Detection is most efficient, for example, when faces are presented in veridical skin-colour tones (Bindemann & Burton, 2009) and when their height-to-width aspect ratio is preserved (Pongakkasira & Bindemann, 2015). These characteristics set detection apart from other tasks with faces, such as recognition, which appears unaffected by geometric distortions of height-to-width ratios (Bindemann, Burton, Leuthold, & Schweinberger, 2008; Hole, George, Eaves, & Rasek, 2002) and colour removal (Kemp, White, Pike, & Musselman, 1996; Yip & Sinha, 2002).

However, as the study of face detection in psychology has gained momentum, a number of different approaches have emerged to study this process. One approach is to present highly simplified stimulus displays, which comprise of a single cropped face on a blank screen. This approach is dominant in studies examining face perception with neuroscience methods such as EEG and fMRI, where additional visual context can produce noise in the data or induce artefacts (e.g., De Lissa, McArthur, Hawelka, Palermo, Mahajan, Degno, & Hutzler, 2019; Gao, Xu, Zhang, Zhao, Harel, & Bentin, 2009; Looser, Guntupalli, & Wheatley, 2013; Moulson, Balas, Nelson, & Sinha, 2011). This approach provides a highly-controlled scenario for studying detection. However, in this setup, the problem of face detection may be solved by the mode of presentation itself, as such tasks do not require the localisation of faces across the visual field and their discrimination from competing non-face stimuli. It is therefore unclear whether this situation provides a good proxy for face detection outside of the laboratory or gives rise to distinctly different results.

A different solution is to present faces in photographs of visual scenes (e.g., Bindemann et al., 2010; Bindemann & Burton, 2009; Burton & Bindemann, 2009; Kelly et al., 2019). This approach acknowledges that face detection outside of the laboratory requires a *search* for these targets across the visual field. This search must be able to proceed in many different contexts, which can be variable in terms of their complexity, colour, shape and meaning. By utilising visual scenes for face detection research, it may be possible to draw stronger inferences about the importance of searching for face detection in real-world settings, but the complex visual variation that is inherent in such displays also reduces standardisation across stimulus sets.

A third method to study detection is to embed faces in visual *arrays*, in which a cropped face is presented among a variety of other cropped visual objects (e.g., Hershler & Hochstein, 2009; Langton, Law, Burton, & Schweinberger, 2008; Meissner, Prüfer, Nordt, Semmelmann, & Weigelt, 2018; Ro, Friggel, & Lavie, 2007; Simpson, Maylott, Leonard, Lazo, & Jakobsen, 2019). These studies provide an intermediate solution between face detection with blank displays and scene images, because the immediately surrounding background for faces is also blank, yet other objects are also presented in the visual display. In comparison to experiments with visual scenes, this might facilitate the parsing of faces from the visual background, shifting the emphasis of the detection process onto general shape information such as the head outline.

As each of these three approaches bears different methodological strengths and weaknesses, the scientific inferences that can be drawn about the facial information that supports detection may also vary across these methods. There is some evidence already to support this reasoning. For example, a detection advantage for faces over non-face objects is not observed when these are presented individually on blank background displays in the visual periphery. A detection advantage for faces emerges, however, when the face and object targets are embedded among other stimuli (Hershler, Golan, Bentin & Hoshstein, 2010). This indicates that the information surrounding a face, and hence the methods that are employed to study face detection, can affect the efficiency with which these stimuli are located.

There is also evidence to suggest that different methodological approaches affect *what* is detected. Frontal and profile faces, for example, are detected with equal efficiency when these are presented in isolation at the centre of blank displays, but an advantage for frontal faces emerges when these stimuli are located in visual scenes (Burton & Bindemann, 2009). It is unresolved, however, whether these findings reflect differences in the search for stimuli, which is minimised when faces are presented centrally and more demanding when they can occur across various onscreen locations, or the search for faces on simple versus more complex background displays.

At present, only limited attempts exist to contrast different approaches to face detection, but understanding how these methods affect performance is imperative to further progress in this field. Contrasting detection with blank displays, visual arrays and scenes allows, for example, for a separation of effects that are caused by the search for faces in complex displays (blank displays *versus* arrays and scenes) and the parsing of faces from the image background (blank displays and arrays *versus* scenes). The combination of these methods therefore provides an important route for increasing our understanding of face detection. This study addressed this directly by comparing the detection of faces that are presented in these different display contexts. In each of these tasks, observers were required to detect the presence or absence of a face.

To provide a contrast against which face detection in these conditions could be compared, faces were first shown in a frontal or a profile view. When faces are searched for in visual scenes, profile views are detected less efficiently than frontal views (Burton & Bindemann, 2009). However, this difference is not found for stimuli presented at fixation (Bindemann & Lewis, 2013). If performance for frontal and profile faces is also comparable when these have to be located in blank displays, then the contrast between these face views can give insight into the role that visual search plays in face detection. In turn, if differences between frontal and profile faces arise in blank, array and scene contexts, then this will speak to the importance of display complexity for understanding face detection.

Experiment 4

This experiment examined the detection of frontal and profile faces when these were shown on blank backgrounds, within an array of objects, or embedded in scenes. If these differences in display complexity influence face detection, then performance should decline as complexity increases. Moreover, if the differences between these display conditions speak to the facial information that is useful for detection, then it should take longer to detect profile faces than frontal faces in scenes (see Bindemann & Lewis, 2013; Burton & Bindemann, 2009), whereas these face views should be detected with equal proficiency in blank displays (see Bindemann & Lewis, 2013; Hershler et al., 2010). It is less clear how face view will affect detection in the arrays, which provide an intermediate level of complexity.

Methods

Participants

Forty-three participants completed the experiment (8 male, 33 female, 1 non-binary, 1 undisclosed). A power analysis was conducted with G*Power based on a repeated-measures ANOVA (within-between interaction) with a medium effect size (f = .25, power = .95, number of groups = 1, and number of measurements = 6) and an alpha threshold of $p = .05$, leading to a suggested sample size of 28. We adopted this as a minimum sample size and advertised the experiment online for 90 minutes. The final sample of 43 participants represents all the sign-ups that occurred in this time period.

These participants ranged in age from 18-39 years, with a mean of 24.1 (*SD* = 6.5). Participants were recruited from an online participation website (Prolific) and were paid a small fee to take part. All reported normal or corrected-to-normal vision. All experiments in this study were carried out in accordance with the Declaration of Helsinki for experiments involving human participants. In all experiments, participants provided informed consent to take part.

Stimuli

The stimuli consisted of frontal and profile views of four female faces and four male faces, resulting in a total set of 16 face images. We presented these stimuli without body cues. Variability in clothing and pose makes bodies poor search targets and, even though bodies occupy more space, face detection is not guided by the body (Bindemann et al., 2010). The faces were cropped to remove extraneous background, so that only the internal features of the face, the hair and face outline were visible. Frontal and profile faces were matched in terms of their surface area, by equating the pixel count of these stimuli using the graphics software Photoshop. In addition, 24 images of household objects (e.g., books, clocks, hats) were used as non-face stimuli and were also cropped to remove any background. The same objects were employed in all stimulus displays, but their location was not repeated across the stimulus displays of the array condition. The faces and objects were sized to 2 x 2 cm (76 x 76 pixels at 96 ppi).

In the blank background condition, a single face or object was presented on a white image background subtending 33.9 (w) x 19.1 (h) cm (1280 x 720 pixels at 96 ppi). This background was divided into an invisible 6 (w) x 4 (w) grid of 24 equal-sized stimulus locations. The stimuli were rotated around these locations to create a set of 24 face and 24 non-face displays that served as target-present and target-absent trials for the blank condition.

The same 6 x 4 grid was employed in the array and scene conditions to position the faces, so that target location could be fully counterbalanced within and across conditions. In the array condition, 23 of the non-face objects and one face were used to construct a stimulus array in the same 6 x 4 grid. On face-absent trials, these arrays comprised of the 24 non-face objects. Across all stimulus displays, the objects and faces were also rotated around locations so that the same stimulus never appeared twice in the same location in different arrays. In this way, 24 target-present and 24 target-absent displays were created.

Finally, photographs of 96 scenes depicting settings such as cafes, kitchens and living rooms were employed as stimuli for the scene condition. These scenes were selected so that none displayed people or faces. Half of these scenes served as face-absent trials and did not contain a face. To create face-present scenes, the same faces as in the blank and array conditions were inserted into the remaining scenes, using the 6 x 4 location grid. This also created 24 frontal-face and 24 profile-face displays. An illustration of the blank, array and scene conditions can be viewed in Figure 3.1.

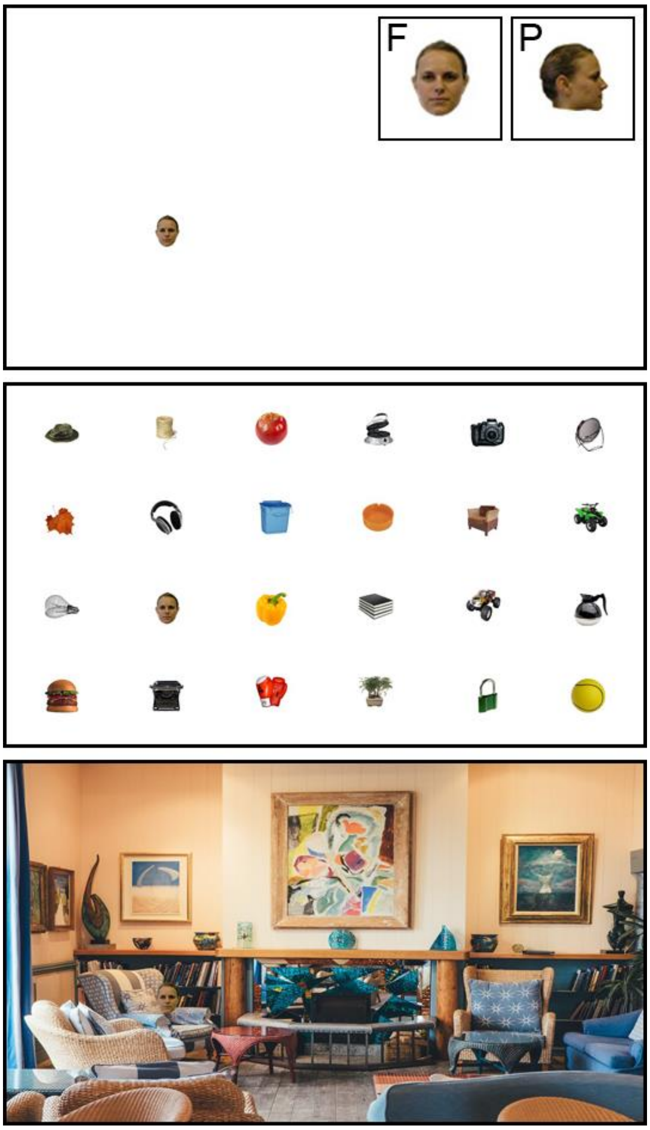


FIGURE 3.1. Example stimuli for the blank (top), array (middle) and scene (bottom) conditions, with insets in the blank display illustrating frontal (F) and profile (P) face targets.

All experiments reported here were created using PsychoPy software (Peirce et al., 2019), and distributed via pavlovia.org for remote (i.e. online) data collection. Participation in these experiments was restricted to particular devices, so that this could only be conducted on a desktop or laptop computer. The onscreen stimulus dimensions were scaled according to the screen size (height) of these devices and height-to-width ratios remained consistent. For example, on a monitor with the dimensions of 30.5 (w) x 20.3 (h) cm, the stimulus displays would appear at a size of 27.3 x 17.2 cm onscreen. Thus, stimulus display size could vary across participants due to differences in device screen, but the size of the stimuli was consistent across conditions within participants.

Procedure

The experiment could only be conducted on a desktop computer or laptop so that the sizing of stimuli was preserved. Each trial began with a fixation cross for 800 milliseconds. This was replaced by a stimulus display, which remained onscreen until a response was registered. Participants were instructed to decide whether a face was present or not in these displays, by pressing ‘F’ on a computer keyboard if they believed a face was present and ‘J’ if they believed there was no face present. The instructions also stated they should respond as quickly and accurately as possible. Participants were informed that faces might appear on a blank background, within an array of objects, or embedded within scenes. However, they were not informed that the faces might appear in frontal or profile view. The three display conditions were administered in blocks of 96 trials, interspersed by short, self-paced breaks. The order of conditions was counterbalanced across participants over the course of the experiment, but the trial order was randomised within blocks for each participant by the display software.

Results

Face present versus face absent

Three participants were excluded from the analysis due to accuracy scores (of 73.6%, 75.7% and 74.0%, combined across conditions and tasks) that fell more than two standard deviations below the sample mean (i.e., scoring below 81.8% accuracy). For the remaining participants, the percentage accuracy scores and the mean correct response times for face-present and face-absent trials were calculated for all conditions. In a first step of the analysis, overall performance on face-present trials and face-absent trials for blank, array and scene displays were analysed to determine whether participants were adhering to task demands. Participants were faster to respond on face-present than absent trials in the blank condition (*M* = 552 ms, *SD* = 83 vs. *M* = 585 ms, *SD* = 84; *t*(39) = 6.37, *p* < .001, *d* = 1.01), the array condition (*M* = 747 ms, *SD* = 120 vs. *M* = 1394 ms, *SD* = 467; *t*(39) = 9.59, *p* < .001, *d* = 1.52), and the scene displays (*M* = 838 ms, *SD* = 128 vs. *M* = 1927 ms, *SD* = 570; *t*(39) = 12.52, *p* < .001, *d* = 1.98), as the search for faces could be terminated once these targets were found.

In addition, participants were less accurate on face-present trials than face-absent trials in the array condition (*M* = 94.06%, *SD* = 6.07 vs. *M* = 99.27%, *SD* = 1.30; *t*(39) = 5.47, *p* < .001, *d* = .87), and the scene condition (*M* = 87.71%, *SD* = 9.04 vs. *M* = 99.01%, *SD* = 1.41; *t*(39) = 7.94, *p* < .001, *d* = 1.26), indicating that they were more likely to miss a face that was present than to detect one where there was none. This pattern was not observed for blank displays (M = 96.20%, SD = 3.92 vs. M = 95.16%, SD = 4.15; $t(39) = 1.71$, $p = .10$, $d = .27$).

Response times

Next, the data of primary interest were analysed, comprising of the median response times for frontal and profile faces in the blank, array and scene conditions. The cross-subject means of these data are summarised in Table 3.1. A 2 (face type: frontal, profile) x 3 (display type: blank, array, scene) repeated-measures ANOVA revealed a main effect of display type, $F(2,78) = 206.45$, $p < .001$, $\eta_p^2 = .84$, and a main effect of face type, $F(1,39) = 44.13$, $p < .001$, $\eta_p^2 = .53$. These effects were qualified by an interaction between these factors, $F(2,78) = 34.52$, $p < .001$, $\eta_p^2 = .47$. Tukey HSD test showed that both frontal and profile faces were detected faster in blank displays than arrays and scenes, all *ps* < .001. In addition, profile faces were detected faster in arrays than in scenes, $p < .001$, whereas response times to frontal faces in arrays and scenes were more similar, $p = .33$. Most importantly, frontal faces were detected more quickly than profile faces in the scene displays, $p < .001$, but not in blank displays, $p = .48$, or stimulus arrays, $p = .25$.

TABLE 3.1. Mean RTs (*ms*), Accuracy (%) and Inverse Efficiency Scores (*IES*) for Frontal (*F*) and Profile (*P*) Faces in the Blank, Array and Scene Conditions in Experiment 1. Parentheses Show the Standard Deviation of the Mean.

	RTs			Accuracy			IES		
	Frontal	*Profile*	*F - P*	*Frontal*	*Profile*	*F - P*	*Frontal*	*Profile*	*F - P*
Blank	559 (93)	545 (80)	14	96.0 (5.42)	96.4 (4.13)	-0.4	584 (100)	566 (82)	18
Array	732 (117)	762 (136)	-30	94.9 (7.15)	93.2 (6.38)	1.7	773 (118)	818 (137)	-45
Scene	768 (132)	908 (146)	-140	91.6 (8.31)	83.9 (11.10)	7.7	847 (167)	1103 (237)	-256

Accuracy

A corresponding ANOVA of response accuracy also revealed main effects of display type, $F(2,78) = 23.21$, $p < .001$, $\eta_p^2 = .37$, and face type, $F(1,39) = 36.22$, $p < .001$, $\eta_p^2 = .48$ and an interaction between factors, $F(2,78) = 14.42$, $p < .001$, $\eta_p^2 = 0.27$. Tukey HSD tests showed that frontal faces were detected with similar accuracy across the blank, array and scene conditions, all $ps \geq .11$. Profile faces were detected more accurately in blank displays than in the array and scene conditions, both $ps < .05$, and detected more accurately in array then scene displays, $p < .001$. These data correspond with response times in that accuracy was higher for frontal faces than profile faces with scene displays, $p < .001$, but not blank displays, $p = 1.00$, or arrays, $p = .51$.

Inverse efficiency scores

In addition, inverse efficiency scores (IES) were calculated to consider speed-accuracy trade-offs, by dividing reaction times by the proportion of correct responses. Higher IES indicate longer detection times when the proportion of errors is taken into account. A repeated-measures ANOVA of these data revealed main effects of display type, $F(2,78) = 173.41$, $p < .001$, $\eta_p^2 = .82$, and face type, $F(1,39) = 67.91$, $p < .001$, $\eta_p^2 = .64$, and an interaction between factors, $F(2,78) = 63.30$, $p < .001$, $\eta_p^2 = 0.62$. Tukey HSD tests showed that both frontal and profile faces were detected more efficiently in blank displays than arrays and scenes, all $ps < .001$, and in arrays than in scenes, $ps < .03$. Moreover, detection of frontal and profile faces was similar in blank displays, $p = .61$, whereas frontal faces were detected more efficiently than profile faces in both arrays and scenes, $ps < .001$.

Discussion

Experiment 4 demonstrates that the search for faces was slower and less accurate in complex scenes than arrays and blank displays. Moreover, the detection of frontal and profile faces was comparable with blank displays and arrays, while a clear detection advantage for frontal over profile face views was found when these stimuli were embedded in scenes. A similar effect was also observed in inverse efficiency scores for arrays, but this was attenuated in accuracy and response times. These findings provide initial evidence that the efficiency of face detection is affected by the type of display that is used. Moreover, the three display types under investigation differed also in what they revealed about the properties of detection. The scene displays show, for example, that frontal faces possess an advantage for fast and accurate detection over profile faces, which could suggest that cognitive templates for face detection might be tuned preferentially to frontal views, whereas the blank displays do not reveal such differences.

Experiment 5

Experiment 5 was conducted to replicate the main findings of Experiment 4, whilst also controlling for differences between the face conditions. Whereas the detection of profile faces was slower than that of frontal faces in scenes in Experiment 4, these face stimuli differed in several characteristics, such as the visibility of internal features (e.g., one eye in profile view versus a pair of eyes in frontal view) and the shape of the head outline. In Experiment 5, the profile faces were therefore replaced with versions of the frontal faces that were rotated by 90°. These rotated stimuli provide the same low-level visual energies, internal features and face outline as their upright counterparts, but in an orientation in which faces are encountered less frequently outside of the laboratory. If cognitive detection templates for face detection are sensitive to these differences, and the display conditions differ in the extent to which this sensitivity can be revealed, then these upright and rotated face conditions should replicate the pattern observed in Experiment 4.

Methods

Participants, Stimuli and Procedure

Thirty undergraduate students from the University of Kent participated in the experiment for course credit. An additional 20 participants were recruited from an online participation website (Prolific) and were paid a small fee to take part. As in Experiment 4, power analysis with G*Power based on a repeated-measures ANOVA (within-between interaction) with a medium effect size (f = .25, power = .95, number of groups = 1, and number of measurements = 6) and an alpha threshold of $p = .05$ suggested a minimum sample size of 28. For consistency with Experiment 4, we continued to advertise the experiment until the same final sample size of 40 was matched. The same approach was adopted in all subsequent experiments reported here. Participants (12 male, 38 female) ranged in age from 18-49 years, with a mean of 25.6 (*SD* = 9.2). All reported normal or corrected-to-normal vision. The stimuli and procedure were identical to Experiment 4, except that the profile face conditions were replaced by copies of the frontal faces that were rotated 90º clockwise. Thus, the upright and rotated face conditions were identical except for the orientation of these targets.

Results

Face present versus face absent

Three participants were excluded from the analysis due to accuracy scores (of 47.6%, 63.2% and 65.6%, combined across conditions and tasks) that fell more than two standard deviations below the sample mean (i.e., scoring below 69.7% accuracy). For the remaining participants, the percentage accuracy scores and the median correct response times for face-present and face-absent trials were calculated for blank, array and scene displays to determine whether participants were adhering to task demands. Participants were faster to respond on face-present than absent trials in the blank condition (*M* = 559 ms, *SD* = 98 vs. *M* = 609 ms, *SD* = 103; *t*(46) = 9.02, *p* < .001, *d* = 1.32), the array condition (*M* = 832 ms, *SD* = 158 vs. *M* = 1518 ms, *SD* = 603; *t*(46) = 9.01, *p* < .001, *d* = 1.31), and the scene displays (*M* = 882 ms, *SD* = 188 vs. *M* = 1967 ms, *SD* = 1103; *t*(46) = 7.32, *p* < .001, *d* = 1.07), as the search for faces could be terminated once these targets were found.

In addition, participants were less accurate on face-present than face-absent trials in the array condition (*M* = 88.79%, *SD* = 11.07 vs. *M* = 98.05%, *SD* = 4.31; *t*(46) = 6.50, *p* < .001, *d* = .95), and the scene condition (*M* = 86.13%, *SD* = 14.28 vs. *M* = 96.68%, *SD* = 7.74; *t*(46) = 6.21, *p* < .001, *d* = .91), indicating that they were more likely to miss a face what was present than to detect one where there was none. This effect was not observed for blank displays (*M* = 95.08%, *SD* = 6.62 vs. *M* = 94.15%, *SD* = 7.90; *t*(46) = 1.22, *p* = .23, *d* = .18).

Response times

The data of primary interest consisted of the median response times for upright and rotated faces in the three display conditions. The cross-subject means of these data are illustrated in Table 3.2. A 2 (face type: upright, rotated) x 3 (display type: blank, array, scene) repeated-measures ANOVA showed a main effect of face type, $F(1,46) = 79.03$, $p < .001$, $\eta_p^2 = .63$, a main effect of display type, $F(2,92) = 135.71$, $p < .001$, $\eta_p^2 = .75$ and an interaction between these factors, $F(2,92) = 21.39$, $p < .001$, $\eta_p^2 = .32$. Tukey HSD test showed that both upright and rotated faces were detected faster in blank displays than in arrays and scenes, both *ps* < .001, and as quickly in arrays as in scenes, *ps* > .25. Importantly, upright faces were detected faster than rotated faces in arrays, $p < .001$, and scenes, $p < .001$, whereas detection of upright and rotated faces was comparable with blank displays, $p = .99$.

TABLE 3.2. Mean RTs (*ms*), Accuracy (%) and IES for Upright (*U*) and Rotated (*R*) Faces in the Blank, Array and Scene Conditions in Experiment 2. Parentheses Show the Standard Deviation of the Mean.

	RTs			Accuracy			IES		
	Upright	*Rotated*	*U - R*	*Upright*	*Rotated*	*U - R*	*Upright*	*Rotated*	*U - R*
Blank	556 (105)	562 (102)	-6	95.8 (7.37)	94.3 (6.92)	1.5	583 (111)	596 (105)	-13
Array	775 (158)	889 (180)	-114	92.9 (9.75)	84.7 (13.41)	8.2	842 (183)	1083 (308)	-241
Scene	811 (202)	954 (196)	-143	89.2 (12.0)	83.1 (17.64)	6.1	922 (246)	1214 (446)	-292

Accuracy

For accuracy, a 2 (face type) x 3 (display) ANOVA also revealed main effects of face type, $F(1,46) = 52.32$, $p < .001$, $\eta_p^2 = .53$, and display type, $F(2,92) = 19.68$, $p < .001$, $\eta_p^2 = .30$, and an interaction between these factors, $F(2,92) = 10.34$, $p < .001$, $\eta_p^2 = .18$. Tukey HSD tests showed that rotated faces were detected more accurately in blank displays than in arrays and scenes, *ps* < .001, but as quickly in arrays as in scenes, $p = .91$. Upright faces were also detected more accurately in blank displays than in scenes, $p = .007$, but accuracy was similar for blank displays compared with arrays, *p* = .41, and for arrays compared with scenes, *p* = .02. However, upright faces were detected more accurately than rotated faces in arrays, *p* < .001, and in scenes, *p* = .001, but detection of upright and rotated faces was comparable in blank displays, *p* = .40.

Inverse efficiency scores

IES also revealed main effects of display type, *F*(2,92) = 110.87, *p* < .001, η_p^2 = .71, and face type, *F*(1,46) = 50.45, *p* < .001, η_p^2 = .52, and an interaction between factors, *F*(2,92) = 19.84, *p* < .001, η_p^2 = 0.30. Tukey HSD test showed that both upright and rotated faces were detected more efficiently in blank displays than arrays and scenes, all *ps* < .001. Upright faces were also detected faster in arrays than scenes, *p* = .01, but this was not found for rotated faces, *p* = .17. More importantly, the detection of upright and rotated faces was comparable in blank displays, *p* = .92, whereas upright faces were detected faster than rotated faces in both arrays and scenes, *ps* < .001.

Discussion

This experiment replicated the important aspects of Experiment 4 with upright and rotated faces. The longest response times and lowest face detection accuracy were again observed in the scene condition in comparison to arrays and blank displays, indicating that the search for faces becomes more challenging in more complex displays. Once again, these different contexts also affected face detection, whereby upright and rotated faces were detected with equal efficiency in blank displays, but upright faces were detected more effectively in scenes than their rotated counterparts.

The advantage for upright faces was also observed in the array condition of Experiment 5. This differs from Experiment 4, where the corresponding difference between upright and frontal faces was attenuated in accuracy and RTs, and only present when these scores were combined with inverse efficiency scores. This difference could be explained in terms of the parsing of faces from the visual background in blank and array displays compared to scenes. Face detection in arrays and blank displays might predominantly reflect the processing of external face shape information such as head outline, which could be harder to distinguish in complex scene contexts that wrap tightly around the face stimuli. This could lead to comparable detection of frontal and profile faces in arrays but not scenes, where other visual information also has to be used in order to detect faces. However, whereas both frontal and profile faces are encountered routinely in our daily lives, rotated faces represent a more unusual portrayal and may therefore not present external shape information that is as useful as that of upright and profile faces in visual arrays. This might explain why frontal *and* profile views are detected equally well in arrays, whereas frontal and rotated faces are not.

Overall, these results are consistent with the notion that face detection operates differently across different display types and, in turn, that this influences what these paradigms can reveal about the detection process. This experiment indicates that the search for faces in more complex visual displays such as arrays and scenes is sensitive to the orientation of a face.

Experiment 6

Experiment 5 showed that detection is sensitive to the orientation of a face when this requires search in arrays and scenes. Experiment 6 examined further which aspects of a face might drive this effect. The rotated faces that were employed in Experiment 5 differed from upright faces in a number of ways, for example, by changing the height-to-width ratio of the stimuli relative to the upright scene context, which influences face detection (see Pongakkasira & Bindemann, 2015). In order to further explore the role of face shape in detection, and extend the range of stimuli with which to compare upright faces, this experiment examined whether the inversion of faces produces similar effects on detection. This manipulation exerts strong effects on other tasks with faces, such as recognition (Farah, Wilson, Drain, & Tanaka, 1995; Valentine, 1988; Yin, 1969). In contrast, inversion appears to produce mixed results in tasks that require face detection. In simple visual displays, inverted faces appear to be detected as efficiently as upright faces (see Bindemann & Burton, 2008), which might occur because inverted faces also retain the height-to-width ratio of upright faces and some feature arrangements (e.g., a horizontal pair of eyes). On the other hand, detection performance appears to decline with inversion in more complex displays, such as picture grids that are searched for a face target (Lewis & Edmonds, 2003). Here, we ask whether inverted faces are detected as efficiently as upright faces in arrays and scenes, or whether they produce decrements in detection performance similar to those produced by profile and rotated faces.

Methods

Participants, Stimuli and Procedure

Forty-five participants completed the experiment (25 male, 18, female, 1 non-binary, 1 undisclosed). Participants ranged in age from 18-40 years, with a mean of 27.8 (*SD* = 6.0). Participants were recruited from an online participation website (Prolific) and were paid a small fee to take part. All reported normal or corrected-to-normal vision. The stimuli and procedure were identical to previous experiments except that faces rotated by 90º were now replaced with inverted faces (i.e. faces rotated by 180º).

Results

Face present versus face absent

One participant was excluded from the analysis due to accuracy scores of 1.4% (combined across conditions and tasks), which fell more than two standard deviations below the sample mean (i.e., scoring below 64.3% accuracy). For the remaining participants, the percentage accuracy scores and the correct response times for face-present and face-absent trials were calculated for blank, array and scene displays. Once again, participants were faster to respond on face-present than absent trials in blank condition (*M* = 563 ms, *SD* = 87 vs. *M* = 600 ms, *SD* = 87; *t*(43) = 7.29, *p* < .001, *d* = 1.10), the array condition (*M* = 833 ms, *SD* = 131 vs. *M* = 1550 ms, *SD* = 449; *t*(43) = 11.65, *p* < .001, *d* = 1.76), and the scene displays (*M* = 871 ms, *SD* = 140 vs. *M* = 1879 ms, *SD* = 711; *t*(43) = 10.61, *p* < .001, *d* = 1.60).

In addition, participants were less accurate on face-present trials than face-absent trials in the array condition (*M* = 93.27%, *SD* = 8.84 vs. *M* = 99.10%, *SD* = 1.30; *t*(43) = 4.36, *p* < .001, *d* = .66), and the scene condition (*M* = 87.36%, *SD* = 11.56 vs. *M* = 98.77%, *SD* = 1.41; *t*(43) = 6.58, *p* < .001, *d* = .99), but not in blank displays, (*M* = 96.63%, *SD* = 2.78 vs. *M* = 96.26%, *SD* = 3.75; *t*(43) = .61, *p* = .55, *d* = .09).

Response times

The data of primary interest were the response times for upright and inverted faces in the three display conditions. These are shown in Table 3.3. A 2 (face type: upright, inverted) x 3 (display type: blank, array, scene) repeated-measures ANOVA showed main effects of face type, *F*(1,43) = 95.71, $p < .001$, $\eta_p^2 = .69$, and display type, *F*(2,86) = 234.82, $p<.001$, $\eta_p^2 = .85$ and an interaction between these factors, *F*(2,86) = 25.14, $p < .001$, $\eta_p^2 = .37$. Tukey HSD tests showed that upright and inverted faces were detected faster in blank displays than in arrays and scenes, all *ps* <.001. However, there was no difference in detection between arrays and scenes for either of these face types, both *ps* > .06. The comparisons of main interest showed that upright faces were detected more quickly than inverted faces in arrays, $p < .001$, and in scenes, $p < .001$, whereas detection of upright and inverted faces was comparable in blank displays, $p = .31$.

TABLE 3.3. Mean RTs (*ms*), Accuracy (%) and IES for Upright (*U*) and Inverted (*I*) Faces in the Blank, Array and Scene Conditions in Experiment 3. Parentheses Show the Standard Deviation of the Mean.

	RTs			Accuracy			IES		
	Upright	*Inverted*	*U - I*	*Upright*	*Inverted*	*U - I*	*Upright*	*Inverted*	*U - I*
Blank	557 (89)	568 (89)	-11	96.9 (3.38)	96.4 (3.77)	0.5	575 (89)	589 (84)	-14
Array	777 (125)	889 (153)	-112	95.6 (8.54)	91.0 (10.4)	4.6	832 (257)	1004 (314)	-172
Scene	786 (132)	957 (186)	-171	89.7 (9.74)	85.0 (14.6)	4.7	883 (152)	1180 (435)	-297

Accuracy

An analogous ANOVA for response accuracy also showed main effects of face type, *F*(1,43) = 25.45, $p < .001$, $\eta_p^2 = .37$, and display type, *F*(2,86) = 19.78, $p < .001$, $\eta_p^2 = .32$, and an interaction between these factors, *F*(2,86) = 5.03, $p = .009$,

η_p^2 = .10. Tukey HSD tests showed that inverted faces were detected more accurately in blank displays than in arrays and scenes, all *ps* = .03. Upright faces were detected more accurately in blank displays than scenes, *p* < .001, but there was no difference for this face type between blank displays and arrays, *p* = .90. Accuracy for both upright and inverted faces was higher in arrays than scenes, *ps* < .05. Furthermore, the detection accuracy for upright and inverted faces was comparable in blank displays, *p* = .98, but was higher for upright faces than inverted faces in arrays, *p* = .001, and in scenes, *p* = .02.

Inverse efficiency scores

IES also revealed main effects of display type, *F*(2,86) = 83.32, *p* < .001, η_p^2 = .66, and face type, *F*(1,43) = 53.12, *p* < .001, η_p^2 = .55, and an interaction between factors, *F*(2,86) = 18.68, *p* < .001, η_p^2 = .30. Tukey HSD tests showed that both upright and inverted faces were detected more efficiently in blank displays than arrays and scenes, all *ps* < .001. Inverted faces were also detected more efficiently in arrays than scenes, *p* = .004, but this was not found for upright faces, *p* = .50. Consistent with the response time and accuracy data, detection of upright and inverted faces was similar in blank displays, *p* = .35, whereas frontal faces were detected more efficiently than profile faces in both arrays and scenes, *ps* < .001.

Discussion

In Experiment 6, the detection of upright frontal faces was faster and more accurate than that of inverted faces in arrays and scenes, but not in blank displays. These inverted stimuli retain the height-to-width aspect ratio of upright faces, as well as the arrangement of some internal features (e.g., a pair of eyes). Despite this, detection was impaired, implying that upright face detection cannot be based solely on these facial characteristics - if it was, then upright and inverted face detection performance would be equivalent.

These results strengthen the case that presenting faces on blank backgrounds utilises different processes for detection than the search for faces in arrays and scenes. On blank backgrounds, profile faces (Experiment 4), rotated faces (Experiment 5) and inverted faces (Experiment 6) were classified as effectively as frontal faces. This indicates that under these circumstances, performance is driven by general detection processes common to both faces and objects, rather than a mechanism that operates on identifying specific facial properties. In arrays and scenes, on the other hand, differences between these face conditions emerge, pointing to cognitive face detection mechanisms that are tuned preferentially to upright frontal faces.

We now turn to the question of which facial aspects drive these detection effects, by utilising the contrast between the display conditions. In the face perception literature, a distinction is often made between the contribution of internal facial features (such as the eyes, nose and mouth) and external facial features (such as hair and face outline) to face identification (e.g., Ellis, Shepherd, & Davies, 1979; Moscovitch & Moscovitch, 2000; Nachson & Shechory, 2002; Young, Hay, McWeeney, Flude, & Ellis, 1985). In the profile, rotated and inverted face conditions that were employed in Experiments 4 to 6, the internal and external features were manipulated simultaneously. In the next experiment, we examine the contribution of both types of features to detection by manipulating internal and external face information independently.

Experiment 7

In this experiment, the design of the preceding experiments was retained, by contrasting face detection in blank displays, arrays and scenes. However, the internal and external facial features were now manipulated independently. Each of these features contributes to face detection when these are presented individually in search arrays (see Experiment 5 in Hershler & Hochstein, 2005), but here we examined which of these contributes more strongly to face detection by manipulating one set of features whilst preserving the other. This was achieved by rotating the internal features of frontal faces through 90º while the external features remained in an upright orientation, or by rotating the external features while retaining an upright orientation for internal features. The detection of these hybrid-feature faces was compared with upright frontal faces in which the orientation of both internal and external features was preserved. If the internal or external features contribute more strongly to face detection, then the faces in which these features are preserved in their original orientation should perform more similarly to upright faces when these are embedded in arrays and scenes.

Methods

Participants

Forty-six participants completed the experiment (17 male, 26 female, 3 non-binary). Participants' ages ranged from 18-40 years, with a mean of 27.6 (*SD* = 6.4). Participants were recruited via Prolific and were paid a small fee to take part.

Stimuli and Procedure

The stimuli and procedure were identical to the preceding experiments, except for the following changes. As in Experiments 4 to 6, the stimuli for the upright condition consisted of frontal views of four female and four male faces. Each of these faces was then manipulated in graphics software (Adobe Photoshop) so that either the internal or external features were rotated clockwise through 90º, while the other feature set remained in an upright orientation. An example of these stimuli can be viewed in Figure 3.2.

In the experiment, each participant completed 432 trials, comprising of 72 target-present and 72 target-absent trials in the blank, array and scene conditions. These trials were subdivided further for each face condition, into 24 upright face trials, 24 external-upright face trials (i.e., with internal features rotated 90º), and 24 internal-upright trials (i.e., with external features rotated 90º). As in previous experiments, the display conditions were blocked, but the trial order was randomised within blocks. Participants were given a break after each block of 144 trials.

FIGURE 3.2. Illustration of the stimuli for Experiment 7, showing a face that is intact and upright (left), with internal-upright features (middle), and with external-upright features (right).

Results

Face present versus face absent

Two participants were excluded from the analysis due to accuracy scores of 5.1% and 39.4% (combined across conditions and tasks), which fell more than two standard deviations below the sample mean (i.e., scoring below 60.8% accuracy). For the remaining participants, the percentage accuracy scores and the correct response times for face-present and face-absent trials were calculated for blank, array and scene displays to determine whether participants were adhering to task demands. Participants were faster to respond on face-present than absent trials in the array condition (M = 826 ms, SD = 106 vs. M = 1658 ms, SD = 543; $t(43)$ = 10.75, p < .001, d = 1.62), and the scene displays (M = 864 ms, SD = 140 vs. M = 2068 ms, SD = 948; $t(39)$ = 8.89, p < .001, d = 1.34), but not in the blank display condition (M = 601 ms, SD = 92 vs. M = 607 ms, SD = 79; $t(43)$ = .88, p < .38, d = .13).

In addition, participants were less accurate on face-present than face-absent trials in the array condition (M = 94.03%, SD = 8.53 vs. M =98.90%, SD = 3.49;

t(43) = 5.61, *p* < .001, *d* = .85), and the scene condition (*M* = 91.35%, *SD* = 7.28 vs. *M* = 98.96%, *SD* = 1.34; *t*(43) = 7.02, *p* < .001, *d* = 1.06), but not the blank condition, (*M* = 95.96%, *SD* = 6.51 vs. *M* = 97.03%, *SD* = 3.46; *t*(43) = 1.03, *p* = .31, *d* = .16).

Response times

Next, the response times for correct target-present trials were analysed for the experimental conditions (see Table 3.4). A 3 (face type: upright, external-upright, internal-upright) x 3 (display type: blank, array, scene) repeated-measures ANOVA revealed main effects of face type, *F*(2,86) = 95.03, *p* < .001, η_p^2 = .69, and display type, *F*(2,86) = 113.13, *p* < .001, η_p^2 = .72, and an interaction between these factors, *F*(4,172) = 9.71, *p* < .001, η_p^2 = .18. Tukey HSD tests showed that upright, external-upright and internal-upright faces were detected faster in blank displays than in both arrays and scenes, all *ps* < .001. However, there was no difference in detection speed between arrays and scenes in all face type conditions, *ps* ≥ .33. Additionally, upright, internal-upright, and external-upright faces were detected with similar speed in blank displays *ps* ≥ .29. In contrast, both upright faces and external-upright faces were detected faster than internal-upright faces in arrays and scenes, all *ps* < .001, whereas detection was similar for upright faces and external-upright faces in these display conditions, both *ps* = 1.00.

TABLE 3.4. Mean RTs (*ms*), Accuracy (%) and IES for Upright (*U*), External-upright (*E*), and Internal-upright Faces (*I*) in the Blank, Array and Scene Conditions in Experiment 4. Parentheses Show the Standard Deviation of the Mean.

	Upright	*External-Upright*	*Internal-Upright*	*U-E*	*U-I*	*E-I*
			RTs			
Blank	587 (103)	602 (86)	614 (110)	-15	-27	-12
Array	785 (107)	789 (111)	913 (141)	-4	-128	-124
Scene	828 (149)	838 (140)	945 (164)	-10	-117	-107
			Accuracy			
Blank	97.5 (4.61)	95.2 (8.71)	95.2 (8.71)	2.3	2.3	0
Array	96.3 (5.41)	94.4 (4.70)	91.4 (10.1)	1.9	4.9	3
Scene	92.0 (7.56)	95.4 (6.24)	86.6 (10.6)	-3.4	5.4	8.8
			IES			
Blank	602 (98)	639 (128)	658 (201)	-37	-56	-19
Array	815 (127)	892 (506)	1018 (506)	-77	-203	-126
Scene	899 (168)	874 (154)	1100 (238)	25	-201	-226

Accuracy

The accuracy data followed a similar pattern. A 3 (face type) x 3 (display type) ANOVA revealed main effects of face type, $F(2,86) = 20.65$, $p < .001$, $\eta_p^2 = .32$, and display type, $F(2,86) = 6.45$, $p = .002$, $\eta_p^2 = .13$, and an interaction between these factors, $F(4,172) = 10.15$, $p < .001$, $\eta_p^2 = .19$. Tukey HSD tests showed that upright and internal-upright faces were detected more accurately in blank displays than in scenes, *ps* < .007, but not in arrays, *ps* > .14. External-upright faces were detected with similar accuracy levels across all three display conditions, *ps* = 1.00. In addition, accuracy was similar across the three face conditions in blank displays, all *ps* ≥ .62. In the arrays, upright faces were detected more accurately than internal-upright displays, $p = .004$, whereas detection accuracy was similar for the upright and external-upright conditions, $p = .94$, and for the internal-upright and external-upright conditions, $p = .07$. Finally, in the scene conditions, external-upright faces were detected more accurately than upright and internal-upright faces, both *ps* < .007, and upright faces were also detected more accurately than internal-upright faces, $p < .003$.

Inverse efficiency scores

IES also revealed main effects of face type, $F(2,86) = 32.76$, $p < .001$, $\eta_p^2 = .43$, and display type, $F(2,86) = 51.37$, $p < .001$, $\eta_p^2 = .54$, and an interaction between factors, $F(4,172) = 6.72$, $p < .001$, $\eta_p^2 = 0.14$. Tukey HSD tests showed that upright, external-upright and internal-upright faces were detected more efficiently in blank displays than in arrays and scenes, all *ps* < .007, upright faces were detected more efficiently in arrays than scenes, p = .02, where there was no difference between arrays and scenes for internal-upright and internal-upright faces, *ps* ≥ .71.

The primary interest was the comparison of face types for each display condition. The detection of upright, internal-upright, and external-upright faces was similar in blank displays, *ps* ≥ .33, as well as in for internal-upright and external-upright in the array conditions, *ps* ≥ .18. In contrast, both upright faces and external-upright faces were detected faster than internal-upright faces in scenes, all *ps* < .001, and upright faces were detected faster than internal-upright faces in array conditions $p < .001$, whereas detection was similar for upright faces and external-upright faces in arrays and scenes display conditions, $p = .88$.

Discussion

This experiment shows that internal and external features exert distinct effects on face detection in visual arrays and real-world scenes. Detection was delayed and less accurate when the external features of a face were rotated by 90º while the internal features remained upright, whereas a similar decrement in performance was not observed when rotated internal features were presented in the context of upright external features. Moreover, the detection of faces in which the orientation of external features was preserved was as fast and (more) accurate than intact faces, in which internal and external features were presented in their typical arrangement. These results indicate that external facial information is *sufficient* for effective face detection, and that the inclusion of upright internal features does not confer any additional benefits for detection.

Experiment 8

The previous experiments indicate that face detection is modulated by visual context. Blank, array and scene displays vary in visual complexity, and face perception differs across these contexts, indicating that this is a key component of detection. However, there are many differences between these visual displays. This final experiment therefore examines the impact of visual complexity on face detection more systematically, by manipulating the content of the natural scenes using Voronoi tessellation. In this method, also referred to as Thiessen polygons, visual displays are transformed into a honeycomb lattice of smaller cells that summarize colour information across a region of an image. By controlling the level of tessellation, the complexity of an image can be systematically manipulated. In Experiment 8, this technique was employed to determine how complexity affects face detection. For this purpose, we reverted to the upright and inverted face stimuli, as these produced clearer differences across different background conditions than frontal and profile views (c.f., Experiments 4 and 5). The detection of upright and inverted faces was then measured with the intact scenes employed in Experiments 4 to 7, and this was compared with Voronoi versions of these scenes in which complexity is gradually reduced. Based on the results of Experiments 4 to 7, the detection of upright faces was expected to be more efficient than that of their inverted counterparts in intact scenes. This difference should attenuate as the complexity of Voronoi scenes decreases.

Methods

Participants

Forty-one undergraduate students from the University of Kent participated in the experiment for course credit. Participants (8 male, 33 female) ranged in age from 18-38 years, with a mean of 19.97 (SD = 3.38). All reported normal or corrected-to-normal vision.

Stimuli

The 96 naturalistic scenes from the previous experiments were also employed as scene stimuli here, and were sized to the same dimensions of 33.9 (w) x 19.1 (h) cm (1280 x 720 pixels at 96 ppi). These stimuli were then processed with a MATLAB script to create four different levels of Voronoi tessellation to create scenes of low, medium and high visual complexity (see https://uk.mathworks.com/matlabcentral/fileexchange/130299-im2voronoi, using filter settings of 25, 15 and 5). An example of these manipulations is illustrated in Figure 3.3 and shows how scene complexity is reduced gradually across Voronoi levels. The same upright and inverted faces as in Experiment 6 were then placed onto each scene in the same positions as in Experiments 4-7.

Procedure

As in the previous experiment, each trial began with an 800 ms fixation cross, followed by a stimulus display, which would remain onscreen until a response was registered by pressing ‘F’ for face present or ‘J’ for absent. Participants were instructed to respond as quickly and accurately as possible but were not informed about the orientation of the faces.

In this manner, each participant completed a total of 192 trials, comprising of two blocks of 96 trials. In the first block, each scene was only shown once, in either the face-present (48 trials) or face-absent conditions (48 trials). In addition, scene complexity was manipulated systematically within blocks, so that 25% of trials depicted original scenes or high, medium and low complexity Voronoi scenes, respectively. The second block was structured in the same way, except that each scene was presented in a different complexity condition. For example, if a high complexity Voronoi scene was presented in block 1, then the same scene with low complexity Voronoi would be presented in block 2. However, over the course of the experiment, the frequency with which each scene appeared in any of the complexity conditions was counterbalanced across participants. Finally, the trial order was randomised for each participant within each block.

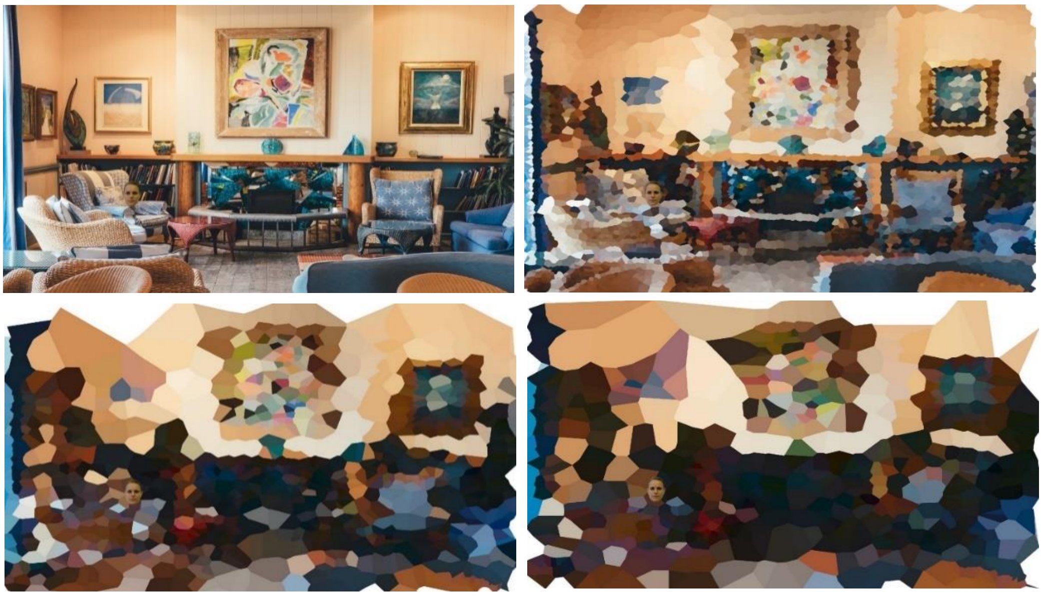


FIGURE 3.3. Illustration of the stimuli for Experiment 8, showing an upright face on an original, unfiltered scene (top left), and scenes of high (top right), medium (bottom left) and low complexity (bottom right).

Results

Face present versus face absent

One participant was excluded from the analysis due to an accuracy score of 53.13% (combined across conditions and tasks) that fell more than two standard deviations below the overall sample mean (i.e., scoring below 77.9% accuracy). For the remaining participants, the percentage accuracy scores and the mean correct response times for face-present and face-absent trials were calculated for all conditions.

In the first step of the analysis, overall performance on face-present trials and face-absent trials for blank, array and scene displays were analysed to determine whether participants were adhering to task demands. Participants were faster to respond on face-present than face-absent trials in the original scenes (*M* = 841 ms, *SD* = 192 vs. *M* = 1496 ms, *SD* = 675; *t*(39) = 6.88, $p < .001$, *d* = 1.09), and across the high (*M* = 731 ms, *SD* = 129 vs. *M* = 1015 ms, *SD* = 296; *t*(39) = 7.90, $p < .001$, *d* = 1.25), medium (*M* = 644 ms, *SD* = 102 vs. *M* = 841 ms, *SD* = 204; *t*(39) = 7.39, $p < .001$, *d* = 1.17), and low visual complexity Voronoi conditions (*M* = 625 ms, *SD* = 116 vs. *M* = 765 ms, *SD* = 171; *t*(39) = 6.97, $p < .001$, *d* = 1.10).

In addition, participants were less accurate on face-present than face-absent trials in the original scenes (*M* = 85.73%, *SD* = 14.18 vs. *M* = 97.29%, *SD* = 4.38; *t*(39) = 5.56, $p < .001$, *d* = .88), and the high (*M* = 85.42%, *SD* = 10.42 vs. *M* = 98.02%, *SD* = 3.28; *t*(39) = 7.45, $p < .001$, *d* = 1.18) and medium complexity Voronoi conditions (*M* = 93.85%, *SD* = 5.74 vs. *M* = 97.40%, *SD* = 3.85; *t*(39) = 4.79, $p < .001$, *d* = .76). In contrast, accuracy was comparable for face-present and face-absent scenes in the low complexity Voronoi condition (*M* = 96.98%, *SD* = 4.62 vs. *M* = 98.33%, *SD* = 2.95; *t*(39) = 1.65, *p* = .11, *d* = .26).

Response times

The median correct response times were analysed next and are displayed in Table 3.5. A 2 (face type: upright, inverted) x 4 (scene complexity: original, high, medium, low) repeated-measures ANOVA of these data revealed an effect of face type, *F*(1,39) = 29.33, $p < .001$, $\eta_p^2 = .43$, and an effect of scene complexity, *F*(3,117) = 58.75, $p < .001$, $\eta_p^2 = .60$. This was qualified by an interaction between face type and Voronoi level, *F*(3,117) = 13.03, $p <.001$, $\eta_p^2 = .25$.

TABLE 3.5. Mean RTs (*ms*), Accuracy (%) and IES for Upright (*U*) and Inverted (*I*) Faces at Original, High, Medium and Low Voronoi scenes in Experiment 5. Parentheses Show the Standard Deviation of the Mean.

	RTs			Accuracy			IES		
	Upright	*Inverted*	*U - I*	*Upright*	*Inverted*	*U-I*	*Upright*	*Inverted*	*U-I*
Original	761 (181)	921 (241)	-160	90.0 (14.52)	81.5 (16.07)	8.5	871 (269)	1184 (450)	-313
High	730 (156)	733 (136)	-3	86.3 (13.15)	84.6 (12.60)	1.7	860 (197)	886 (216)	-26
Medium	620 (97)	668 (117)	-48	94.0 (8.00)	93.8 (6.74)	0.2	664 (116)	716 (139)	-52
Low	620 (106)	630 (139)	-10	97.1 (6.41)	96.9 (5.56)	0.2	640 (110)	652 (147)	-12

Tukey HSD tests showed that both upright and inverted faces were detected more slowly as scene complexity increased across all possible comparisons between conditions, all *ps* < .05, but for three exceptions. Upright faces were detected with similar speed in the original and high complexity scenes, p = .79. Similarly, upright and inverted faces were detected with similar speed in low and medium complexity Voronoi scenes, both *ps* ≥ .39.

Of primary interest were the comparisons between upright and inverted faces at each level of scene complexity. These showed that upright faces were detected faster than inverted faces in the original scenes, p < .001, and in the medium complexity Voronoi scenes, p = .007. In contrast, upright and inverted faces were detected with similar speed in the high and low complexity Voronoi scenes, both *ps* ≥ .99.

Accuracy

A corresponding ANOVA of response accuracy also revealed main effects of face type, $F(1,39) = 11.03$, $p = .002$, $\eta_p^2 = .22$, scene complexity, $F(3,117) = 23.61$, $p < .001$, $\eta_p^2 = .38$, and an interaction between these factors, $F(3,117) = 4.69$, $p = .004$, $\eta_p^2 = .11$. Tukey HSD test showed that accuracy for upright faces was similar across comparisons of scene complexity, all *ps* ≥ .11, except for decreased accuracy in the high than in either the medium or low complexity Voronoi scenes, both *ps* < .05. Similarly, detection accuracy for inverted faces was comparable between the original and high complexity Voronoi scenes, and between the medium and low complexity Voronoi scenes, $p \geq .12$. There was lower detection accuracy with original and high complexity scenes than with either medium or low complexity scenes, all *ps* < .001.

The primary interest again comprised of the comparisons between upright and inverted faces at each level of scene complexity. These showed that upright faces were detected more accurately than inverted faces in the original scenes, $p < .001$, whereas accuracy was comparable for upright and inverted faces in the high, medium and low complexity Voronoi conditions, all *ps* ≥ .48.

Inverse efficiency scores

Consistent with the response time and accuracy data, IES also showed main effects of face type, $F(1,39) = 34.56$, $p < .001$, $\eta_p^2 = .47$, scene complexity, $F(3,117) = 47.56$, $p < .001$, $\eta_p^2 = .55$, and an interaction, $F(3,117) = 19.06$, $p < .001$, $\eta_p^2 = 0.33$. Tukey HSD tests showed that the IES pattern followed response times closely. Thus, both upright and inverted faces were detected less efficiently as scene complexity increased across all possible comparisons between conditions, all *ps* < .001, but for three exceptions. Upright faces were detected with similar efficiency in the original and high-complexity scenes, $p = 1.00$. Similarly, upright and inverted faces were detected with similar efficiency in low and medium-complexity Voronoi scenes, both *ps* ≥ .09.

Direct comparisons between upright and inverted faces showed that upright faces were detected more efficiently than inverted faces in the original scenes, $p < .001$, and medium complexity Voronoi scenes, $p = .02$. In contrast, upright and inverted faces were detected with similar efficiency in high and low complexity Voronoi scenes, both $ps \geq .99$.

Discussion

This experiment shows that faces are located more slowly and less accurately in natural scenes than in Voronoi-filtered images in which scene complexity has been reduced. Importantly, this affects the visual information that is utilised for detection, as upright and inverted faces were detected with similar speed and accuracy in scenes of low complexity. In contrast, upright faces demonstrated a clear detection advantage over inverted faces in more complex scenes. These effects were most pronounced when the original scenes were compared with the least detailed Voronoi images, with intermediate levels of Voronoi tessellation producing more mixed results. These findings converge with the preceding experiments, by demonstrating that face detection becomes more challenging as a function of the complexity of the context within which faces are presented. By studying face detection in more complex stimulus displays, it becomes possible to dissociate the detection of different face stimuli (e.g., upright versus inverted faces). This reveals the facial characteristics that are most useful for detection.

General Discussion

This research shows that face detection draws on different processes depending on the visual context in which faces are viewed. Across five experiments, detection was fastest, most accurate and unaffected by changes in view, rotation and inversion when faces were presented in visual displays that were otherwise blank (Experiments 4 to 7) or of greatly reduced complexity (Experiment 8). These findings converge with other studies that show that face and non-face objects are detected with equal efficiency under similar conditions (Hershler et al., 2010) and indicate that the problem of face detection is solved by the mode of presentation itself in blank displays, whereby stimuli are located quickly irrespective of their appearance.

A different pattern emerged with visual arrays and scenes, both of which gave rise to slower detection performance and more detection errors than blank displays. Most importantly, detection in these displays was consistently sensitive to differences in facial information, with frontal faces outperforming profile faces (Experiment 4), and upright faces outperforming faces that were rotated through 90º or inverted (Experiments 5, 6 and 8). These findings provide evidence that the context in which faces are presented not only affects the efficiency of the detection process, but that it can also provide insight into the stimulus characteristics that are important for detection.

In Experiment 7, we explored this directly by investigating whether the internal or external features are particularly important for face detection in complex contexts. Faces in which the internal features were rotated by 90º while the external features remained upright were detected as quickly, and even more accurately, than intact upright frontal faces. In contrast, detection performance declined when internal features remained upright and external features were rotated. The differences between these conditions indicate that cognitive templates for face detection are not tuned strongly to the internal features of faces, but rely more strongly on information such as the head outline and general face-shape. This explanation converges with other studies that demonstrate that detection is impaired by geometric distortions of face shape (Pongakkasira & Bindemann, 2015), when faces are presented partially (Burton & Bindemann, 2009), or in unnatural colour tones (Bindemann & Burton, 2009; Prunty et al., 2022).

This reasoning gains further traction when considering that the challenge of face detection is to find faces that appear in the visual periphery, outside of foveal vision, where faces appear to have a particularly strong detection advantage (Hershler et al., 2010). The detail of internal facial features may be difficult to resolve with the loss of acuity in the human visual field that such eccentric presentations entail (see, e.g., Burton, Bindemann, Langton, Schweinger, & Jenkins, 2009; Rousselet, Husk, Bennett, & Sekuler, 2005). This should shift the emphasis of cognitive detection templates towards information that remains accessible even under lowered resolution, such as a general face shape.

This reasoning could also be reconciled with the detection disadvantage for profile faces in Experiment 4 (see also Bindemann & Lewis, 2013; Burton & Bindemann, 2009), which provide different shape information due to the intrusion of the hair region in this view and a different face outline. This could also explain why *similar* detection performance was obtained for internal and external face regions in other studies (see Hershler & Hochstein, 2005), where the faces in both of these conditions were cropped to preserve elliptical face shapes. The current study extends these findings by demonstrating that the efficiency of face detection is maintained when general face-shape information is preserved even when the typical arrangement of internal features is not.

Considering that face detection was impaired for profile faces (Experiment 4) and frontal faces that were rotated by 90º (Experiment 5) in comparison to upright frontal face views, the question arises of whether symmetry is an important element for optimising face detection. In the inverted face conditions of Experiment 6 and 8, symmetry information was retained but detection was attenuated compared to upright faces. Other studies have also shown that the detection advantage for frontal faces over other face views persists when symmetry is eliminated by presenting only one half of a face (Burton & Bindemann, 2009). This indicates that symmetry *per se* is not responsible for the detection advantage for upright frontal faces. Perhaps this information must be combined with other facial cues to optimise detection performance. However, such an account must also explain the effects of scene complexity that were observed consistently across all five experiments here.

One explanation for these findings could be that in blank or very simple visual displays (such as the low complexity Voronoi scenes), a face could be differentiated from non-face objects (or the visual background) on just a single visual feature, for example, such as skin-colour *or* face outline. This would avoid the need to process these stimuli in greater depth, resulting in fast and accurate detection performance – as was observed across all experiments here. In contrast, the parsing of faces from more complex backgrounds might require a combination of information, such as colour *and* shape, as non-face objects might share some of these individual features. Such an account would resonate with studies with non-face stimuli that have demonstrated how the search for single-feature targets becomes more challenging when these are embedded in more varied or complex visual displays (e.g., Duncan & Humphreys, 1989; Santhi & Reeves, 2004; Treisman, 1991). It also resonates with theories in which conjunctions of features can guide search for a target more effectively than single features (e.g., Wolfe et al., 1989). In the face domain, such an account would be consistent with reports that several sources of information are important for face detection in scenes

(Bindemann & Burton, 2009; Burton & Bindemann, 2009; Pongakkasira & Bindemann, 2015; Prunty et al., 2022), and that representations of faces in which different sources of information are combined are detected more effectively than separable facial features (Hershler & Hochstein, 2005).

Finally, we note that the current findings converge with other studies which suggest that detection may be quite distinct from other tasks with faces (see Bindemann & Lewis, 2013; Pongakkasira & Bindemann, 2015; Qarooni, Prunty, Bindemann, & Jenkins, 2022). The recognition of facial identity and emotion, for example, rely on internal facial features to function optimally (see, e.g., Ellis, Shepherd, & Davies, 1979; Toseeb, Keeble, & Bryant, 2012; Wegrzyn, Vogt, Kireclioglu, Schneider, & Kissler, 2017), whereas the detection of faces appears unaffected by changes to the internal features (such as rotation in Experiment 7). This emphasises the importance of understanding the process of face detection in its own right. This study extends these observations by demonstrating that *how* detection is investigated determines *what* can be learned about this process. In contrast to other tasks with faces, such as recognition and emotion perception, detection should be studied with complex visual displays that necessitate the search for faces for a fuller understanding of this process.

Chapter 4:

The Role of Features in Detecting and Classifying Faces

Introduction

Chapter 3 investigated whether face detection depends on the context in which these stimuli are seen, by presenting faces in blank displays or embedding these in arrays of objects or images of scenes. To examine the influence of context, different face conditions, which are known to disrupt detection performance, were compared. Thus, detection for frontal faces was compared with faces in profile view (Experiment 4), faces rotated by 90° (Experiment 5), and turned upside-down (Experiment 6).

Several insights emerged from these experiments. First, context consistently affected face detection performance, whereby search times increased and accuracy decreased with increasing display complexity. Therefore, detection was most difficult with scenes and best with blank displays, with arrays typically producing an intermediate level of performance. A final experiment then applied Voronoi transformations to scenes to confirm that the complexity of stimulus displays modulates face detection (Experiment 8). Detection decreased with the increase of Voronoi complexity, therefore confirming that differences in search efficiency for faces can be attributed to display complexity.

Second, the orientation of faces consistently affected detection performance, whereby frontal faces were located more effectively than their profile, rotated and upside-down counterparts. However, these effects were only observed with arrays and scenes, whereas performance was comparable across these different face conditions in blank displays and Voronoi displays with the lowest complexity (Experiments 4, 5, 6 and 8). A further experiment examined which facial characteristics drive these detection effects, by rotating either the internal or external facial features by 90° while their counterparts remained upright

(Experiment 7). Faces with rotated internal features were detected as efficiently as their intact frontal faces, whereas detection was impaired when external features were rotated. Taken together, these experiments demonstrate that context influences what can be learned about the face detection process. In complex visual arrays and scenes, detection proceeds more effectively when external facial features are preserved in an upright orientation, whereas the internal configuration of features appears to be of lesser importance. These findings are consistent with a cognitive detection template that focuses on general face-shape information (see Pongakkasira & Bindemann, 2015).

In the current chapter, these effects are examined further to confirm the facial characteristics that are important for detection. A constraint of the experiments in Chapter 3 is that detection performance was always measured *indirectly*, by recording observers’ response times and accuracy. This approach is consistent with other studies in this domain (e.g., Bindemann & Burton, 2009; Burton & Bindemann, 2009; Bindemann & Lewis, 2013), but cannot reveal the locus at which specific visual information becomes important for detection or whether the effects reported in Chapter 3 occur during the detection stages of face processing at all. For example, it is possible that profile faces are found as quickly as frontal faces in scenes (in Experiment 4) but, once these stimuli are located, observers require additional time to confirm that a looked-at profile view is, in fact, a face. Eye-tracking allows for more precise measures of face processing, as the point at which a saccade is initiated, when a face is fixated, and how long it is fixated for, can be determined, providing more clarity as to the stages of detection.

In this chapter, a new gaze-contingent paradigm is therefore introduced that provides a more direct measure of the visual characteristics that drive detection *during* the localisation of a face. In this paradigm, participants' eye movements are tracked during the search for faces in scenes. The trajectory of their eye movements (i.e., saccades) is then used to manipulate the stimulus display immediately *after* a face has been located but *before* a detection decision is registered. This manipulation of the stimulus display focuses specifically on the face content, by either removing faces or by changing the identity of faces after they have been detected by participants, but before they can be fixated directly. These manipulations are contrasted with a control condition in which no changes to the display occur.

This gaze-contingent paradigm is based on several premises. First, viewers' eye movements dictate what is occurring within the display, by changing the appearance of face targets in milliseconds as viewers saccade towards these stimuli. This method can create changes to faces just outside of foveal vision, and as the acquisition of visual information is paused during saccades (Houtmans & Sanders, 1983), these changes should not be perceived. Second, visual acuity is highest for stimuli presented in the centre of the visual field (the fovea) and declines in the periphery (Westheimer, 1965; Kalloniatis & Luu, 2007). Consequently, the visual information that can be perceived in the periphery differs from central vision. By manipulating the presentation of faces as they cross from peripheral into central vision during an eye movement, this gives insight as to what is drawing the initial saccade when a face is detected in the *visual periphery*, and then how responses are made once the change has occurred, and *central vision* takes priority as the target is fixated.

The question of primary interest here is whether the visual information that drives face detection does, in fact, vary across the visual field. Peripheral vision seems to hold importance in detection processes, such as the search for faces, as it is relied upon for viewing complex scenes. For example, when search displays are masked in the periphery, inspection times increase, indicating a reliance on peripheral vision for scene processing (Van Diepan & d'Ydewalle, 2003). Searching for non-face targets such as objects or letters in real-world scenes is also optimal when peripheral vision is available and decreases when only foveal vision is accessible (Nuthmann & Canas-Bajo, 2022). Peripheral vision is therefore vital for the spatial search of targets and the processing of complex displays. In turn, as face detection involves the localisation of faces in the wider visual environment, observers must be able to perceive low-level information (e.g., colour and shape; see Bindemann & Burton, 2009; Burton & Bindemann, 2009; Prunty et al., 2022) that is important for this process in the periphery, and this may differ from more detailed information in faces that can be extracted at fixation.

However, the existing evidence is mixed in this regard. There is evidence, for example, that detection declines as faces are presented at greater eccentricities in the visual field (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005). Additionally, brain responses (such as the N170 component, and activity in the Fusiform Face Area) are stronger when faces are presented centrally compared to the periphery (De Lissa et al., 2019; Levy et al., 2001). This indicates that face detection might rely on detail-specific processing that deteriorates in the periphery rather than low-level visual information. On the other hand, faces hold a detection advantage over other stimuli in the visual periphery. For example, observers are more likely to saccade towards faces than objects when these are presented at opposite ends of the visual field (Crouzet, Kirchner & Thorpe, 2010; Little, Jenkins & Susilo, 2021), and faces retain this advantage over objects even at extreme eccentricities of 80^o in the visual field and in ‘crowded’ conditions (Boucart et al, 2016). This indicates that it is important to dissociate processing in the periphery and at fixation to understand face detection, but still leaves open the question of which facial characteristics drive this process.

In this chapter, this is investigated across a series of four experiments. To first determine whether changes to faces are perceived in peripheral vision, Experiment 9 uses the gaze-contingent paradigm to either remove faces, change their identity or make no changes. If changes to faces are not noticed, then changes in identity should not impact performance as observers shift from peripheral view to fixation, as this factor is irrelevant to face detection (i.e., this process must operate on *any* face, regardless of who it belongs to), whereas removal of a face should cause a decline in performance at the point at which a detection decision is made. This is examined by measuring the accuracy and speed with which faces are detected across conditions and the time in which these stimuli are fixated in displays. In Experiment 10, the presence of internal facial features is then manipulated to determine their importance in the detection of faces. If these features do not drive detection, then their removal or addition should not impact performance. Yet if such features are important, then faces with the internal features initially removed should cause a decline in search performance in the periphery, but increased performance at fixation. Alternatively, faces in which the features are removed at fixation should demonstrate the same fixation rates as intact faces, but potentially poorer performance than when features appear at fixation. Finally, Experiments 11 and 12 examine which facial information observers can perceive in the periphery when instructed explicitly to do so. This provides an important comparison to determine whether any gaze-contingent changes in Experiments 9 and 10 affect face detection because visual information from internal features is not *seen* or is not *used*.

Experiment 9

This experiment examined whether detection is influenced by changing facial information in the visual periphery. Eye-tracking was utilised to make changes to the visual display and to examine fixation on the face targets. As observers made a saccade towards a target face, it was either removed from the scene or exchanged for the face of another person. By contrasting this with a condition in which no changes are made, it becomes possible to determine whether the detection process is sensitive to visual information such as identity or relies on more general characteristics that are shared across different faces.

Methods

Participants

Sixty participants completed the experiment (5 male, 54 female, 1 non-binary). A power analysis was conducted with G*Power based on a repeated-measures ANOVA (within factors) with a medium effect size (f = .25, power = .95, number of groups = 1, and number of measurements = 3) and an alpha threshold of $p = .05$, leading to a suggested sample size of 43. This was adopted as the minimum sample size, and advertised the experiment online for 4 weeks, with a maximum sign-up of 60. The final sample of 60 participants represents all sign-ups that occurred in this time period.

These participants ranged in age from 18-36, with a mean of 19.5 (*SD* = 2.38). Participants were undergraduate students at the University of Kent who participated for course credit. All reported normal or corrected-to-normal vision. All participants signed a written informed consent form to confirm taking part, which occurred in all following experiments.

Stimuli & Materials

The stimuli consisted of eight full frontal faces, comprising of four males and four females. These faces were cropped to remove any extraneous background, with only the face and hair visible, and sized to a height of 53 pixels at a resolution of 72ppi, with some variation in width due to slight differences in face and hair shape. These faces were presented on a set of 72 scenes, which depicted settings such as living rooms, cafes and restaurants, and were sized to 1024 (w) x 768 (h) pixels at a resolution of 72 ppi. Faces would appear in one of six locations in the periphery of the screen. The locations were determined by a 5 (w) x 3 (h) grid, in which faces would appear in the centre of the top, middle and bottom locations to the left and right sides of the grid. Only the outer locations were used to ensure faces remained far enough out in the peripheral vision for the manipulations to be effective.

FIGURE 4.1. Example stimuli for scenes, in which faces stayed the same (Top image would remain), changed identity (Top image would switch to the middle image) and disappeared (Top image would switch to the bottom image).

The set of 72 scenes were divided into three conditions, corresponding to *change*, *disappear* and *no-change* trials. On change trials, as participants saccaded towards a face, it switched to a different identity. The exchanged faces were matched on general dimensions, such as gender and hair colour. On disappear trials, the face was removed from the scenes as participants saccaded towards it. These visual changes were triggered when saccade trajectories hit a boundary that fell on a face region that spanned approximately 170 pixels away from the face within the scene. The size of these regions were calculated using visual angles. The eye-tracking chin rest was positioned approximately 69 cm away from the display screen and central vision is considered to be around 3-5 degrees. Therefore, changes were triggered by eye movements that moved within 5 degrees of the outer height and width dimensions of the faces (corresponding to approximately 6 cm or 170 pixels). Finally, in the no-change condition, the stimulus display remained the same, with the same face on the same background, regardless of participants' eye movements (see Figure 4.1). In addition, 72 face-absent scenes were also created, in which no faces were present. Of these scenes, 24 were from a new stimulus set, and 48 were repeated from the face-present conditions.

Six regions of interest (ROI) were also created, which covered the placements of the faces in their locations, in order to perform the eye-tracking analysis. These consisted of black ovals on a white background, with a width and height of 100 x 125 pixels (see Figure 4.2). This size was used to account for the variations in the widths of the face, to allow for slight differences in placement (some faces were moved slightly to blend in with the scenes) and so that fixations that fell just outside the face were also included.

Figure 4.2. Example of a face presented in a scene in the top left corner (left image) and the corresponding ROI that covers the area of the face (right image)

Procedure

The stimuli were displayed using SR-Research Experiment Builder Software (version 1.1.0), with the dimensions of 1024 x 768 pixels, on a 21inch monitor that was connected to an EyeLink 1000 eye-tracking system, running at a 1,000 Hz sampling rate. Participants' right eye was tracked, unless calibration issues arose, in which tracking was switched to the left eye. Head movements were stabilised using a chinrest located 69cm from the monitor. Participants were calibrated using nine fixation dots on the display monitor, which was then repeated to validate the calibration using the standard EyeLink procedure. Calibration was repeated until the validation indicated accurate measurement. Calibration occurred at the beginning of the experiment, and every 36 trials.

Each trial started with a fixation dot, which was used for drift correction. This dot appeared in the centre of the screen and appeared in between adjacent trials. This also ensures that each participant starts each trial looking at the centre of the screen. Once the participant fixated on the dot, the trial could begin.

Participants were told to indicate whether a face was present or absent in the scene by pressing the ‘f’ key for face present and the ‘j’ key for face absent. The instructions asked that they respond as quickly and accurately as possible. Participants were not informed of the changes that would occur.

Each trial was displayed until a response was registered. In the change and disappear conditions, displays would switch with millisecond timing once the participants’ saccades hit the 170px boundary, and the second display would remain until the participant made a response. Trials were presented in four blocks of 36, to give participants breaks and to re-calibrate. Each participant took part in all 144 trials, comprising of 24 trials in the change condition, 24 trials in the disappear condition, 24 trials in the no-change condition, and 72 face-absent trials. The trial order was randomised for each participant.

After the experiment, a questionnaire was also administered to determine the extent to which participants were aware of the changes. The first 2 questions remained open questions as to not prime participants with the answers of interest. These questions were as follows: “Did you notice anything unusual during the experiment? If so, please describe it.”, “Did you notice anything unusual about the faces during the experiment? If so, please describe it?”. These were then followed by 3 more specific questions: “In any of the scenes, did you see faces suddenly disappear?”, “In any of the scenes, did you see any of the faces change appearance while you were trying to look at them?” and “In any of the scenes, did you see any faces suddenly appear where there had been none a brief moment ago?”. Participants gave written answers to each of the questions and answered each in the order presented.

Results

Detection Times and Accuracy

Five participants were excluded from the analysis due to accuracy scores that fell more than two standard deviations below the sample mean (i.e., scoring below 77.8% accuracy) or reaction times that fell more than two standard deviations above the sample mean (i.e., responses above 3314 ms), leaving a sample of 55 participants for analysis. For the remaining participants, the percentage accuracy scores and the mean correct response times for face-present and face-absent trials were calculated for all conditions. For the purpose of this experiment, the disappear condition was considered a face-present trial, with the correct response indicating that a face was present, to determine if these faces could clearly be identified in peripheral vision. In the first step of the analysis, overall performance on face-present trials (i.e., collapsed across face type) and face-absent trials were analysed to determine whether participants were adhering to task demands. Participants were faster to respond on face-present trials (M = 1169 ms, SD = 215) than in face-absent trials (M = 2334 ms, SD = 927), $t(53) = 10.66$, $p < .001$, $d = 1.45$, as the search for faces ends once these stimuli are found. Participants were also less accurate on face-present trials (M = 81.06%, SD = 8.75) than on face-absent trials (M = 97.63%, SD = 3.40), $t(54) = 12.06$, $p < .001$, $d = 1.63$, indicating that faces were more likely to be missed than falsely detected.

TABLE 4.1. Accuracy (%), Reaction time (*ms*), IES, Trials Fixated (%), First fixations (*ms*), and Fixation durations (*ms*) for No-Change (*N*) Changed (*C*) and Disappearing (*D*). Parentheses Show the Standard Deviation of the Mean.

	No Change	*Change*	*Disappear*	*N - C*	*N - D*	*C - D*
Accuracy	91.6 (7.71)	94.5 (5.38)	57.1 (23.23)	-2.9	34.5	37.4
RTs	1173 (210)	1155 (234)	1197 (301)	18	-25	-42
IES	1278 (187)	1220 (219)	3346 (4435)	58	-2068	-2126
Trials Fixated	45.0 (27.39)	45.6 (30.95)	16.0 (19.43)	-0.6	29	29.6
First Fixations	800 (243)	773 (193)	573 (179)	27	227	200
Fixation Duration	232 (47)	246 (57)	305 (175)	-14	-73	-59

The mean response times for the experimental conditions (change, disappear, no-change) were then analysed and are provided in Table 4.1. For reaction times, one participant had missing data due to an error in the recording of the data, so was additionally excluded from this analysis. A repeated-measures ANOVA revealed no effect of experimental condition on reaction times, $F(2,106) = 1.58$, $p = .21$, $\eta_p^2 = .03$. In contrast, a corresponding ANOVA of response accuracy revealed an effect of condition, $F(2,108) = 118.58$, $p < .001$, $\eta_p^2 = .69$. Tukey HSD tests showed that accuracy decreased when faces disappeared compared to when no-change to faces occurred during detection, $p < .001$, and when faces changed identity, $p < .001$. In addition, faces were detected more accurately in the change condition than in the no-change condition, $p = .009$.

To investigate possible speed-accuracy trade-offs, inverse efficiency scores (IES) were also calculated by dividing reaction times by the proportion of correct responses. Therefore, higher IES indicate less efficient search times when the proportion of errors is considered. Again, one participant had missing data and was therefore excluded from this analysis. A repeated-measures ANOVA revealed an effect of condition, $F(2,106) = 12.32$, $p < .001$, $\eta_p^2 = .19$. Tukey HSD tests showed that responses were less efficient for faces that disappeared than the change and no-change conditions, *ps* < .003. In addition, responses were also more efficient when faces changed compared to the no-change condition, $p = .02$.

Eye-Tracking Data

In the next step of the analysis, participants' eye movements were examined to determine the speed with which faces were first fixated (as a measure of detection), and the duration of the first fixation (as a measure of face classification). Eye movements were preprocessed by integrating fixations of less than 80 ms with the preceding or following fixation if that fixation lay within half a degree of visual angle (see Rayner & Pollatsek, 1989). The eye fixation data were then compared with a set of regions of interest (ROIs) which reflected the location of faces in the scenes.

The face regions were fixated on 35.78% of all trials, corresponding to 44.96% (SD = 27.39) of no-change trials, 45.56% (SD = 30.95) of change trials, and 16.00% (SD = 19.43) of disappear trials. Note that the lower percentage of fixations for face disappear trials is expected, due to the disappearance of these faces during a saccade towards the target ROI. A repeated-measures ANOVA revealed a difference in the number of trials fixated and face type, $F(2,106) = 72.73$, $p < .001$, $\eta_p^2 = .58$. Tukey HSD tests showed that the face regions of no-change and change trials were more likely to be fixated than on disappear trials, *ps* < .001. No difference was found in the percentage of fixations between change or no-change faces, $p = .97$.

The time to the first fixation on the face region then was calculated to determine whether experimental condition influenced fixation time. These data are provided in Table 4.1. A repeated-measures ANOVA revealed an effect of condition on first fixation times, $F(2,80) = 22.59$, $p <.001$, $\eta_p^2 = .36$. Tukey HSD tests showed that areas where the faces had disappeared were fixated faster than faces that changed or when no changes occurred, both $ps < .001$. However, there was no difference in fixation times in the no-change and change conditions, $p = .40$, which indicates that changes in face identity did not affect detection speed.

Next, the fixation duration was calculated for each condition. A repeated-measures ANOVA revealed an effect of condition on fixation duration, $F(2,80) = 16.06$, $p < .001$, $\eta_p^2 = .29$. Tukey HSD tests showed that disappearing faces regions were fixated for longer than faces in the change condition, $p =.002$, and the no-change condition, $p < .001$. No difference in fixation duration was found between changing and not-changing faces, $p = .13$.

Questionnaire Analysis

The questionnaire was distributed to participants after the experiment to determine their awareness of the changes in their peripheral vision. Open questions were coded as to whether they mentioned faces disappearing, changing or both. When given the open question “Did you notice anything unusual during the experiment?”, only 9% had noticed both that the faces were disappearing and changing. A further 65% had noticed the faces disappear, but only an additional 2% mentioned only the faces changing in appearance. In contrast, when prompted directly with the questions “In any of the scenes, did you see faces suddenly disappear?” and “Did you see any of the faces change appearance while you were trying to look at them?”, 100% of participants responded with yes to faces disappearing, but only 29% reported seeing faces change. In addition, 38% of participants also falsely believed that faces would suddenly appear. Finally, 20% of participants had spontaneously reported seeing “flashing” faces.

Discussion

As expected for Experiment 9, detection was impaired by the disappearance of faces. This reflects the uncertainty that this condition creates, whereby faces are perceived in the visual periphery and trigger eye movements to their locations, but are no longer visible to observers once their location has moved into foveal view. Crucially, however, detection was not impaired by a change in identity. If anything, faces in the change condition were detected more accurately than faces in the no-change condition. This finding indicates that a change in identity is not detrimental to detection, presumably because detection utilises peripheral vision and therefore *precedes* fixation of the faces – and therefore also precedes the change in identity. This reasoning is also supported by the fixation times, which demonstrates that faces were fixated with comparable speed in both the change and no-change conditions.

The experiment provides additional information that observers were not sensitive to facial identity in the periphery, to yield converging evidence that this does not influence the detection process. Firstly, fixation durations on the face regions were substantially longer in the disappear than the change and no-change conditions, which suggests that observers had some awareness that a face stimulus had been perceived but was no longer present. In contrast, the fixation durations were comparable for the change and no-change conditions, indicating again that a change in identity information was not registered in the periphery. Second, participants were much more likely to report explicitly that faces had disappeared (74% of observers) during the experiment than that they had noticed changes in identity (11%).

Taken together, these findings provide evidence that face processing should be dissociated in the periphery and at fixation to understand the process of detection, and that this can provide insight into the facial characteristics that drive this process. In this experiment, the removal of a face but not a change in identity impaired detection, indicating that the latter information is irrelevant to the task.

Experiment 10

Experiment 9 indicates that information to distinguish different faces, such as identity, does not impact detection performance. This is consistent with a process that must operate on *any* face, regardless of who it belongs to. However, the question remains as to whether internal features that are shared across faces, such as the general configuration of a pair of eyes above a central nose and mouth, are important for detection. In Chapter 3, face detection was impaired when internal facial features were preserved but the external features were rotated. In contrast, detection proceeded unhindered when internal features were rotated and external features were preserved. This raises the possibility that detection does not rely on either a full face or the internal features despite their importance for other face tasks, such as the recognition of identity and emotion (see, e.g., Ellis, Shepherd, & Davies, 1979; Toseeb, Keeble, & Bryant, 2012; Wegrzyn, Vogt, Kireclioglu, Schneider, & Kissler, 2017). However, the experiments relied on indirect measures of detection such as response times and accuracy.

In the current experiment, the relevance of internal features for face detection is therefore examined directly using the gaze-contingent paradigm introduced in Experiment 9. Thus, the internal features (e.g., eyes, nose, mouth) were either removed as observers saccaded towards a target, to leave a shape-colour face template with intact external features (e.g., only head shape, skin colour and hair); or the internal features were added as observes saccaded towards the target, so that the stimulus changed from an external featured face to an intact face. By comparing these conditions with the detection of intact faces in which no change occurred, the importance of internal and external features *during* the detection process was assessed. If a basic colour-shape template without internal features such as the eyes, nose and mouth is the basis for face detection, as is suggested by the results of Chapter 3, then there should be no difference between the intact face condition and the initial presentation of faces without internal features. In contrast, if these internal features support detection, then this process should be impaired when this information is initially absent from the search displays.

Methods

Participants

Sixty participants completed the experiment (8 male, 50 female, 1 gender fluid, 1 non-binary). These participants ranged in age from 18-45, with a mean age of 20.52 years (*SD* = 4.79). Participants were undergraduate students at the

University of Kent, who participated for course credit. All reported normal or corrected-to-normal vision.

Stimuli & Procedure

The stimuli and procedure remained the same as in the previous experiment except for the following changes. The same face stimuli were used, but each face was manipulated in graphics software (Adobe Photoshop) so that the internal features were removed and replaced with just skin tones to create featureless faces. In 24 trials, the full faces would switch to the featureless faces in the *internal-disappear* condition as participants made a saccade towards them. In another 24 trials, the featureless faces would switch to full faces in the *internal-appear* condition. Finally, in another 24 trials, the full faces would stay the same in the *no-change* condition (see Figure 4.3). Participants were instructed to search for a “human face or head” to remove any confusion on seeing a blank featureless face, as some may only categorise a face as a face if internal features are present.

FIGURE 4.3. An example of faces where internal features are intact (top) and where the internal features are removed (bottom). In the no-change condition, the top image remains on screen. In the internal-disappear trials, the top presentation would appear first, followed by the display on the bottom when the boundary was hit. For the internal-appear trials, the display presentation was reversed (bottom to top).

Results

Detection time and accuracy

Four participants were excluded from the analysis due to accuracy scores that fell more than two standard deviations below the sample mean (i.e., scoring below 87.1% accuracy) or reaction times that fell more than two standard deviations above the sample mean (i.e., responses above 3746 ms) leaving a final sample of 56 participants. In the first step of the analysis, overall performance on face-present trials (i.e., collapsed across the experimental conditions) and face-absent trials were analysed to determine whether participants were adhering to task demands. Participants were faster to respond on face-present trials (M = 1199 ms, SD = 212) than in face-absent trials (M = 2848 ms, SD = 1194), $t(55) = 11.45$, $p < .001$, d = 1.53, as the search for faces ends once found. Participants were also less accurate on face-present trials (M = 94.12%, SD = 5.22) than face-absent trials (M = 98.26%, SD = 2.88), $t(55) = 4.95$, $p < .001$, $d = .66$, indicating that faces are more likely to be missed than falsely detected.

In the next step of the analysis, the mean response times were compared for the experimental conditions (internal-appear, internal-disappear, no-change). The mean response times are provided in Table 4.2. A repeated measures ANOVA revealed an effect of condition on reaction times, $F(2,110) = 6.44$, $p = .002$, $\eta_p^2 = .11$. Tukey HSD tests showed the internal-appear condition had no difference in response times compared to the disappear and no-change conditions, both $ps \geq .09$. However, responses were slower in the internal-disappear than the no-change condition, $p = .001$.

A corresponding ANOVA of accuracy also revealed an effect of condition, $F(2,110) = 9.55$, $p < .001$, $\eta_p^2 = .15$. Tukey HSD tests showed that accuracy declined in the internal-disappear trials compared to both the no-change and internal-appear conditions, $ps \leq .01$. No difference in accuracy occurred between the internal-appear and the no-change conditions, $p = .73$.

TABLE 4.2. Accuracy (%), Reaction Times (*ms*), IES, Trials Fixated (*%*), First fixations (*ms*) and Fixation durations (*ms*) for No-Change (*N*), Internal-Appear (*A*) and Internal-Disappear (*D*) faces. Parentheses Show the Standard Deviation of the Mean.

	No Change	*Internal Appear*	*Internal Disappear*	*N - A*	*N - D*	*A - D*
Accuracy	95.4 (5.77)	96.0 (4.77)	91.0 (10.62)	-0.6	4.4	5
RTs	1166 (188)	1194 (257)	1238 (239)	-28	-72	-44
IES	1225 (195)	1244 (256)	1394 (423)	-19	-169	-150
Trials Fixated	51.1 (28.16)	43.1 (34.30)	46.0 (29.01)	8	5.1	-2.9
First Fixations	804 (175)	733 (172)	813 (201)	71	-9	-80
Fixation Duration	228 (55)	225 (56)	240 (64)	3	-12	-15

To further the analysis, inverse efficiency scores (IES) were calculated to consider the trade-off between accuracy and reaction times. A repeated-measures ANOVA revealed a difference in face type, $F(2,110) = 10.39$, $p < .001$, $\eta_p^2 = .16$. Tukey HSD tests showed responses were less efficient in the internal-disappear condition compared to the no-change and internal-appear conditions, both $ps \leq .006$. No differences were found between the no-change and internal-appear trials, $p = .70$.

Eye-tracking data

In the next step of the analysis, the participants' eye-movement data were examined to compare to the behavioural results. Eye movements were preprocessed by integrating fixations of less than 80 ms with the preceding or following fixation if that fixation lay within half a degree of visual angle, or otherwise excluded. The eye fixation data were compared with a set of ROIs where the faces appeared in the scenes.

Overall, the face regions were fixated on 46.74% of trials, corresponding to 51.07% (SD = 28.16) of no-change trials, 43.14% (SD = 34.30) of internal-appear trials, and 46.00% (SD = 29.01) of internal disappear trials. A repeated-measures ANOVA of these data revealed a main effect of face type, $F(2,110) = 8.06$, $p < .001$, $\eta_p^2 = .13$. Tukey HSD tests showed that faces where no-change occurred were more likely to be fixated than faces where the internal features appeared, $p < .001$. No difference was found in the percentage of fixations between internal-disappear faces and internal-appear faces, or no-change faces, all $ps \geq .10$.

The time to the first fixation on the face was calculated next to determine whether the experimental conditions influenced how quickly faces were located in the scene displays. A repeated-measures ANOVA revealed an effect of condition on first fixation times, $F(2,110) = 3.92$, $p = .02$, $\eta_p^2 = .07$. Tukey HSD tests showed that participants were slower to fixate on faces in the internal-disappear condition, compared to when internal features appeared during a saccade, $p = .03$. Faces were also fixated faster in the internal-appear than the no-change condition, $p = .04$, while there was no difference in first fixations between the internal-disappear and no-change condition, $p = .97$.

The duration of the first fixation on the face regions was also calculated for each condition, as a measure of the decision process to decide that these regions did, in fact, contain a face. A repeated measures ANOVA revealed no effect of condition on fixation duration, $F(2,110) = 2.60$, $p = .08$, $\eta_p^2 = .05$.

Questionnaire Analysis

The questionnaire was distributed to participants after the experiment to determine their awareness of the changes to the faces. When given the open question “Did you notice anything unusual during the experiment?” 85% reported not noticing anything, and only 15% mentioned anything to do with the faces changing. However, when prompted with the questions “In any of the scenes, did you see faces suddenly disappear?”, “did you see any of the faces change appearance while you were trying to look at them?” and “Did you see any faces suddenly appear where there had been none a brief moment ago?”, 67% of participants responded with yes to faces disappearing, 79% reported seeing faces change and 24% reported seeing faces appearing.

Discussion

In this experiment, the relevance of internal features for face detection was examined using the gaze-contingent paradigm introduced in Experiment 9. Thus, the internal features (e.g., eyes, nose, mouth) were either removed as observers saccaded towards a face to leave a shape-coloured face template with only external features on display, or a shape-coloured face template without internal features was presented, to which the internal features were added as observers saccaded towards the face. By comparing these conditions with the detection of intact faces in which no such change occurred, the importance of internal and external features *during* the detection process was assessed.

The results demonstrate that the removal of internal features whilst observers are saccading towards a face reduces the speed and accuracy with which these stimuli are classified. In contrast, faces in which the internal features were initially omitted were fixated faster than faces that did not change, or faces that were initially intact and features were thereafter removed. This contrast indicates that two different processes are involved in face detection, and each of these relies on different facial information. The fixation data indicate that internal features are not important for the *initial localisation* of a face. This finding appears to make good sense considering the loss of visual acuity in the periphery (see, e.g., Westheimer, 1965; Kalloniatis & Luu, 2007), which would likely reduce the perception of facial detail, such as the eyes, nose and mouth, in the stimuli presented. Indeed, the current results indicate that this localisation process might even be hampered by the presence of internal features, as their presence reduced the speed with which faces were fixated. The second process benefits from the presence of internal features and appears to take place closer to fixation. This process might reflect a decision-making process that a located candidate region in a scene is, in fact, a face. As this process draws on information provided by foveal vision, it can utilise additional visual detail, such as internal facial features, to make an accurate decision.

Experiment 11

The previous experiment indicates that internal features are not necessary for locating faces in the visual periphery, but are important for deciding that a looked-at face candidate region is, in fact, a face. This finding raises the question of whether internal facial features can be processed in the periphery at all in order to drive detection. In Experiment 11, this is examined directly with a visual discrimination paradigm in which observers are presented with pairs of faces, split across the left and right visual field. One of these stimuli comprises of an intact face, which is paired either with a featureless face with the internal features removed, or a face in which these features are rotated by 90°. The difficulty of this task is manipulated in two ways, both of which are important for face detection. First, the eccentricity of these stimuli in the periphery is manipulated by presenting these faces either at 5° or 10° from fixation (as presented in Rousselet et al., 2005), to establish whether sensitivity to featural information is reduced at greater distances. Second, by presenting the faces on blank displays or embedded in scene displays, it is also possible to compare the influence of display complexity on face processing.

The main question of interest here is whether participants can distinguish between the face conditions by determining the side of the visual field in which the intact face is presented. If observers can distinguish between intact faces and faces without internal features in the periphery, then this would suggest that such information is available to the visual system during face detection, but is not required to facilitate this process (as in Experiment 10). Moreover, if observers can determine the orientation of these features, then this would indicate that this information can be processed in considerable detail. Thus, this experiment will reveal the extent to which observers are sensitive to (i) the presence and (ii) the organisation of facial features in the visual periphery.

Methods

Participants

Sixty participants completed the experiment (5 male, 54 female, 1 non-binary). These participants ranged in age from 18-45 years, with a mean of 19.90 (SD = 4.06). Participants were undergraduate students at the University of Kent who participated for course credit. All reported normal or corrected-to-normal vision.

Stimuli

The stimuli consisted of 24 faces in total. Eight faces (four male and four female) were used as full-frontal faces. The same eight faces were manipulated in graphics software (Adobe Photoshop) so that the internal features were removed, and replaced with just skin tone to create featureless faces, and were also manipulated so that the internal features were rotated 90^o. As previously, the faces were cropped so that there was no background, with only the face and hair being visible (see Figure 4.4). The eye-tracking chin rest was always positioned approximately 69 cm away from the display screen, to give a consistent visual angle. Faces were sized so that the height was 3.0^o of the visual field, and approximately 1.8° wide, with some variation due to differences in face shape.

Faces were presented on either side of the fixation point, either 5° or 10° from the centre. The gaze continency was manipulated so that if participants looked away from the fixation cross, the faces would disappear. This was to ensure that participants were only relying on their peripheral vision to process the faces. The internally rotated or featureless faces were always paired with a full face with matched identity. Each face appeared on both the left and right sides of the fixation point, making 16 trials per condition (see Figure 4.4).

Faces were also presented in this way on both blank displays and within scenes. The scenes consisted of 64 different images of settings such as living rooms, cafes, restaurants, etc. Scenes were sized to 1024 (w) x 768 (h) pixels at a resolution of 72 ppi, with the faces overlayed on top (See Figure 4.4).

Procedure

Participants were given both verbal and written instructions. They were instructed to indicate whether the full face appeared on either the left (by pressing the 'f' key) or right (by pressing the 'j' key) side of the fixation point. Participants were notified of the nature of the other faces and told that if they looked away from the fixation cross, then the faces would disappear. Participants were informed to respond as quickly and accurately as possible. Each trial began with a fixation dot for drift checking, and when fixated, the display would appear. The display would remain until the participant made a response. Trials were presented in 2 blocks of 64, either being presented with the blank background trials first or the scene trials.

This was counterbalanced across participants. Each participant took part in all 128 trials, and the trial order within each block was randomised for each participant.

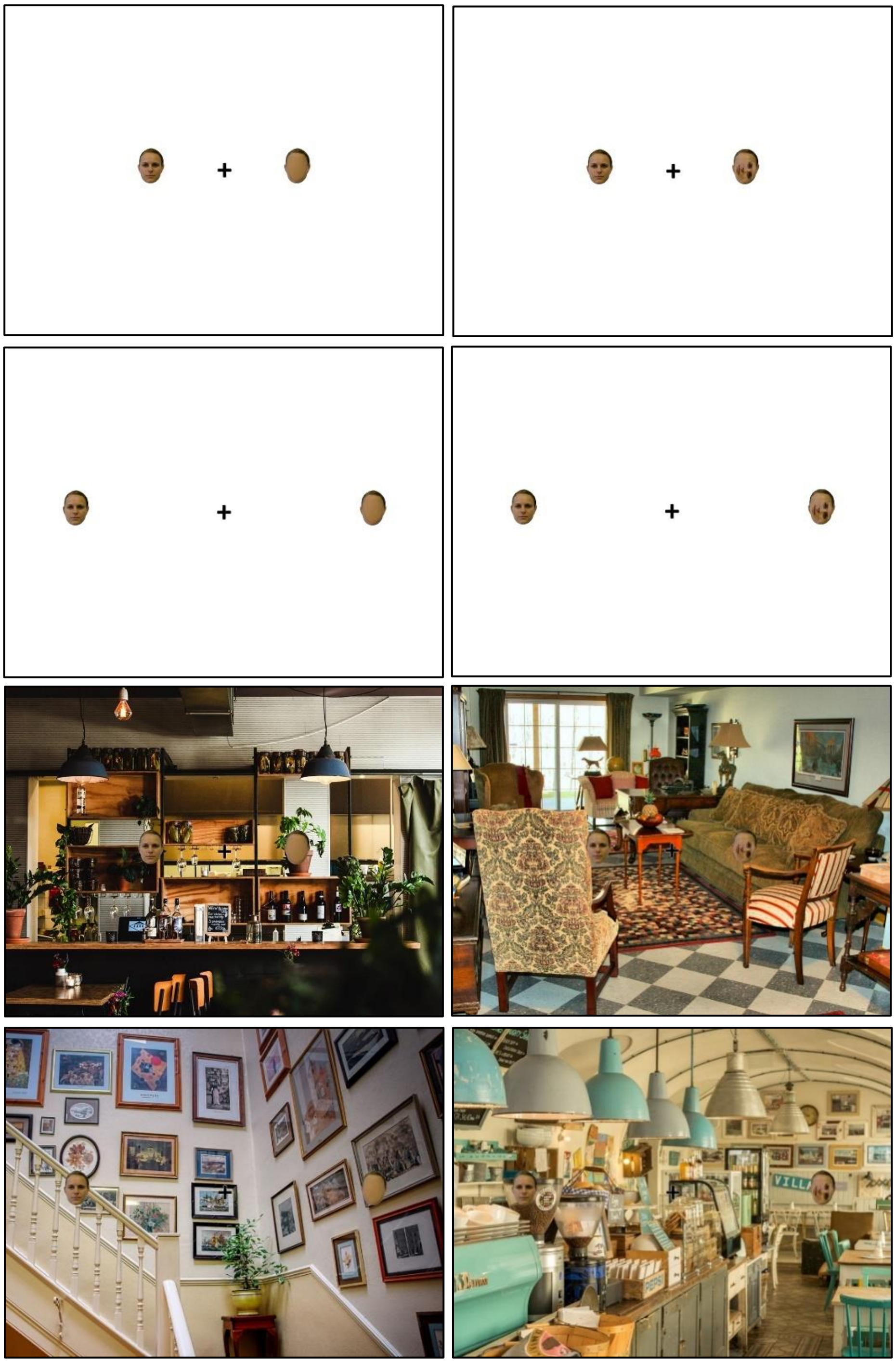


FIGURE 4.4. Example stimuli of the featureless face condition (left) and internally rotated face condition (right) presented at 5º (first and third rows) and 10º (second and fourth rows) and on blank (first and second rows) and scene (third and fourth rows) displays.

Results

Six participants were excluded from the analysis due to either accuracy scores that fell more than two standard deviations below the sample mean (i.e., scoring below 77.80% accuracy) or reaction times that fell more than two standard deviations above the sample mean (i.e., responses above 945 ms), leaving a final sample of 54.

Response Times

First, the percentage accuracy scores and the mean response times were calculated for all conditions, and are presented in Table 4.3. A 2 (face type: featureless, rotated) x 2 (display type: blank, scene) x 2 (eccentricity: 5°, 10°) repeated-measures ANOVA was run to determine the effect on reaction times. This revealed a main effect of face type, $F(1,53) = 171.88$, $p < .001$, $\eta_p^2 = .76$, and an effect of eccentricity, $F(1,53) = 141.51$, $p < .001$, $\eta_p^2 = .73$. These effects were qualified by an interaction between face type and eccentricity, $F(1,53) = 73.03$, $p < .001$, $\eta_p^2 = .58$. At both distances of 5° and 10°, responses were slower for rotated than featureless faces, *ps* < .001. In addition, responses on the featureless and rotated conditions were slower when the faces were presented at 10° than at 5° eccentricity, *ps* < .001.

An effect of display type was also found, indicating that face discrimination was faster when faces were presented on blank backgrounds than scenes, $F(1,53) = 23.45$, $p < .001$, $\eta_p^2 = .31$. There was no interaction between eccentricity and display type, $F(1,53) = 3.02$, $p = .09$, $\eta_p^2 = .05$, between face type and display type, $F(1,53)$

= 1.37, p = .25, η_p^2 = .03, or a three-way interaction, F(1,53) = .03, p = .86, η_p^2 = .00.

TABLE 4.3. RTs (*ms*), Accuracy (%) and IES for Featureless (*F*) and Rotated (*R*) Faces in the Blank and Scene Displays, at both 5° and 10° distances in Experiment 3. Parentheses Show the Standard Deviation of the Mean.

		RTs			Accuracy			IES		
		Featureless	*Rotated*	*F - R*	*Featureless*	*Rotated*	*F - R*	*Featureless*	*Rotated*	*F - R*
Blank	5°	530 (102)	595 (113)	-65	99.2 (2.44)	97.9 (5.43)	1.3	535 (103)	610 (123)	-75
	10°	574 (90)	786 (211)	-212	93.3 (1.65)	76.3 (13.21)	17	616 (97)	1047 (275)	-431
Scene	5°	573 (119)	658 (150)	-85	99.7 (1.45)	96.8 (6.73)	2.9	576 (119)	688 (189)	-112
	10°	653 (130)	880 (204)	-227	99.2 (2.73)	78.9 (12.87)	20	660 (133)	1140 (325)	-480

Accuracy

A corresponding ANOVA of response accuracy also revealed main effects of face type, F(1,53) = 143.49, $p < .001$, η_p^2 = .73, eccentricity, F(1,53) = 213.16, $p <.001$, η_p^2 = .80, and display type, F(1,53) = 12.46, $p < .001$, η_p^2 = .19. These effects were qualified by an interaction between eccentricity and face type, F(1,53) = 123.26, $p < .001$, η_p^2 = .70. At both 5° and 10° eccentricity, featureless faces were distinguished more accurately than rotated faces, p = .05 and $p < .001$, respectively. Both featureless and rotated faces were also distinguished more accurately at 5° than at 10°, $ps < .001$.

There was an additional interaction between eccentricity and display type, F(1,53) = 20.17, $p < .001$, η_p^2 = .28. Faces were discriminated more accurately in both blank and scene displays at 5° than 10° eccentricity, $ps < .001$. At 10°, faces were classified more accurately in scene displays than in blank displays, $p < .001$, however, this was not found at 5°, $p = .77$.

A final interaction was found between face type and display type, $F(1,53) = 5.07$, $p = .03$, $\eta_p^2 = .09$. Featureless faces were distinguished more accurately in scene displays than blank displays, $p < .001$, however, this was not found for rotated faces, $p = .89$. In both blank and scene displays, featureless faces were compared more accurately than rotated faces, *ps* < .001.

Inverse Efficiency Scores

To examine for possible speed-accuracy trade-offs between accuracy and reaction times, inverse efficiency scores (IES) were also calculated. A repeated-measures ANOVA revealed an effect of face type $F(1,53) = 216.84$, $p < .001$, $\eta_p^2 = .80$, an effect of eccentricity, $F(1,53) = 240.14$, $p < .001$, $\eta_p^2 = .82$, and an effect of display type, $F(1,53) = 14.36$, $p < .001$, $\eta_p^2 = .21$. This was qualified with an interaction between face type and eccentricity, $F(1,53) = 157.15$, $p < .001$, $\eta_p^2 = .75$. Both featureless and rotated faces were distinguished more efficiently at 5° than 10°, *ps* < .001. In addition, featureless faces were discriminated more efficiently than rotated faces at both 5° and 10°, *ps* < .001.

An interaction was also found between face type and display type, $F(1,53) = 4.53$, $p = .04$, $\eta_p^2 = .08$. Both featureless and rotated faces were compared more efficiently on blank backgrounds than on scenes, *ps* = .006. Featureless faces were distinguished more efficiently than rotated faces on both blank displays and scene displays, *ps* < .001. No interaction was found between eccentricity and display type,

$F(1,53) = .16$, $p = .69$, $\eta_p{}^2 = .003$, or between all three factors, $F(1,53) = .09$, $p = .77$, $\eta_p{}^2 = .002$.

Discussion

This experiment revealed that observers can distinguish the presence and absence of internal features for faces presented in the periphery. This task becomes more challenging as the eccentricity of faces increases from 5° and 10°. Overall, however, accuracy was high (> 93%) and response times were fast (< 660 ms). A different picture emerged when internal features were rotated by 90°, which increased response times and lowered accuracy to classify faces substantially, to between 76% and 79% across the rotated conditions at 5° and 10° eccentricity (with a chance level of 50%). Taken together, these findings indicate that observers are sensitive to the *presence* but not the *organisation* of facial features in the visual periphery.

These findings have several implications. Since observers can distinguish between intact faces and faces without internal features in the periphery, this suggests that such information is available at some level to the visual system during face detection. However, since the localisation of faces is not facilitated by the presence of internal features, as is evident in Experiment 10, the detection process does not appear to depend critically on this information. This could be explained by the second key finding of this experiment, that observers are not sensitive to the orientation of internal features in the periphery. If the orientation of a basic face schema (e.g., a pair of eyes above a central nose and mouth) cannot be perceived accurately in the periphery, then this suggests that these internal features only provide an imprecise detection cue. Indeed, if observers cannot reliably distinguish upright from rotated internal features, then the presence of such features could be potentially misleading, for example, by drawing observers to face-like candidate regions in the visual field that might capture *some* aspects of internal facial features at different rotations, rather than actual face targets. In that case, it would be beneficial to operate with cognitive templates for face detection that do not rely strongly on such featural information. This could explain why the localisation process was hampered by the presence of internal features in Experiment 10, where their presence reduced the speed with which faces were fixated. This issue is considered further in the General Discussion.

Experiment 12

Experiment 11 demonstrates that faces in which the internal features have been removed could be distinguished with accuracy and speed from intact upright faces in the periphery, whereas faces in which the internal features are rotated are difficult to distinguish from their intact upright counterparts. However, the experimental design allowed for unlimited viewing time of stimuli, which may have enhanced accuracy in this task by allowing attention shifts to the target stimuli to be made. In Experiment 12, this possibility is restricted by limiting the display time of the experimental displays to just 200 milliseconds, to determine the reliability of these findings.

Methods

Participants

Sixty participants completed the experiment (8 male, 50 female, 1 gender fluid, 1 non-binary). These participants ranged in age from 18-38, with a mean of 20.10 (SD = 3.56). Participants were undergraduate students at the University of Kent who participated for course credit. All reported normal or corrected-to-normal vision.

Stimuli & Procedure

The stimuli and procedure remained the same as in the previous experiment except for the following changes. Faces were presented in the periphery for 200ms before disappearing. This timeframe is based on the time it takes to initiate a saccade (Purves et al., 2001), meaning that participants should not have time to shift their attention to either face. The participants would then indicate which side they thought the whole face appeared on before moving on to the next trial.

Results

Four participants were excluded from the analysis due to accuracy scores that fell more than two standard deviations below the sample mean (i.e., scoring below 69.85% accuracy) or reaction times that fell more than two standard deviations above the sample mean (i.e., responses above 951 ms), leaving a final sample of 56.

Response Times

The percentage accuracy scores and the mean response times were calculated for all conditions, and are presented in Table 4.4. A 2 (face type: featureless, internally rotated) x 2 (display type: blank, scene) x 2 (eccentricity: 5°, 10°) repeated-measures ANOVA was conducted to determine the effect of these factors on reaction times. This revealed main effects of face type, $F(1,55) = 303.79$, $p < .001$, $\eta_p^2 = .85$, eccentricity, $F(1,55) = 114.79$, $p < .001$, $\eta_p^2 = .68$, and display type, $F(1,55) = 26.27$, $p < .001$, $\eta_p^2 = .32$. These effects were qualified by an interaction between eccentricity and face type, $F(1,55) = 39.40$, $p < .001$, $\eta_p^2 = .42$. At both 5° and 10° eccentricity, featureless faces were classified faster than rotated faces, *ps* < .001. In addition, response times were faster in both the featureless and the rotated conditions at 5° than at 10° eccentricity, *ps* < .001.

An interaction between eccentricity and display type was also identified, $F(1,55) = 11.06$, $p = .002$, $\eta_p^2 = .17$. At both 5° and 10° eccentricity, faces were classified faster in blank than in scene displays, *ps* < .001. For both blank displays and scene displays, faces were classified faster at 5° than at 10° eccentricity, *ps* < .001. Finally, no interaction between face type and display type, $F(1,55) = .64$, $p = .43$, $\eta_p^2 = .01$, or between all three factors was found, $F(1,55) = .01$, $p = .91$, $\eta_p^2 = .00$.

TABLE 4.4. RTs (*ms*), Accuracy (%) and IES for Featureless and Rotated Faces in the Blank and Scene Displays, at both 5° and 10° distances in Experiment 4. Parentheses Show the Standard Deviation of the Mean.

		RTs			Accuracy			IES		
		Featureless	*Rotated*	*F - R*	*Featureless*	*Rotated*	*F - R*	*Featureless*	*Rotated*	*F - R*
Blank	5°	527 (105)	624 (111)	-97	99.1 (2.51)	94.2 (8.58)	4.9	533 (106)	666 (123)	-133
	10°	545 (101)	727 (170)	-182	93.4 (2.51)	71.9 (13.59)	22	584 (108)	1056 (355)	-472
Scene	5°	584 (118)	693 (164)	-109	98.4 (2.98)	90.0 (12.48)	8.4	594 (119)	791 (253)	-197
	10°	641 (146)	839 (257)	-198	96.9 (5.59)	66.7 (10.73)	30	666 (161)	1294 (468)	-628

Accuracy

A corresponding ANOVA of accuracy also revealed main effects of face type, $F(1,55) = 271.50$, $p < .001$, $\eta_p^2 = .83$, eccentricity, $F(1,55) = 361.41$, $p < .001$, $\eta_p^2 = .87$, and display type, $F(1,55) = 5.33$, $p = .03$, $\eta_p^2 = .09$. These effects were qualified by an interaction between eccentricity and face type, $F(1,55) = 185.73$, $p < .001$, $\eta_p^2 = .77$. At both 5° and 10° eccentricity, accuracy was higher in the featureless than the rotated condition, *ps* < .001. In addition, both featureless and rotated faces were classified more accurately at 5° than at 10° eccentricity, *ps* < .001.

An interaction between face type and display was also found, $F(1,55) = 18.62$, $p < .001$, $\eta_p^2 = .25$. Featureless faces were compared more accurately in scenes than in blank displays, $p = .03$, whereas rotated faces were distinguished more accurately in blank than in scene displays, $p = .005$. However, in both blank and scene displays, featureless faces were distinguished more accurately than rotated faces, *ps* < .001. Finally, no interaction was found between eccentricity and display type, $F(1,55) = 1.69$, $p = .20$, $\eta_p^2 = .03$, or between all three factors, $F(1,55) = 3.34$, $p = .07$, $\eta_p^2 = .06$.

Inverse Efficiency Scores

Once again, inverse efficiency scores (IES) were also calculated to consider speed-accuracy trade-offs. A repeated-measures ANOVA of these data revealed main effects of face type, $F(1,55) = 314.80$, $p < .001$, $\eta_p^2 = .85$, eccentricity, $F(1,55) = 189.95$, $p < .001$, $\eta_p^2 = .78$, and display type, $F(1,55) = 23.96$, $p < .001$, $\eta_p^2 = .30$. This was qualified by an interaction between face type and eccentricity, $F(1,55) = 128.95$, $p < .001$, $\eta_p^2 = .70$. Both featureless and rotated faces were distinguished more efficiently at 5° than 10° eccentricity, *ps* < .001. In addition, featureless faces were also distinguished more efficiently than rotated faces at both 5° and 10° eccentricity, *ps* < .001.

An interaction was also found between face and display type, $F(1,55) = 8.45$, $p = .005$, $\eta_p^2 = .13$. Both featureless and internally rotated faces were classified more efficiently on blank backgrounds than on scenes, *ps* < .001, but featureless faces were also distinguished more efficiently than rotated faces in both types of displays, *ps* < .001.

An additional interaction was also found between eccentricity and display type, $F(1,55) = 4.50$, $p = .04$, $\eta_p^2 = .08$, whereby faces were distinguished more efficiently on blank backgrounds than scenes, both at 5° and 10° eccentricity, both *ps* < .001. Moreover, faces were also distinguished more efficiently at 5° than 10° eccentricity, both in blank and scene displays, *ps* < .001. Finally, a three-way interaction was not found, $F(1,55) = 2.20$, $p = .14$, $\eta_p^2 = .04$.

Discussion

This experiment replicates the main findings of Experiment 11, by demonstrating that observers can distinguish the presence and absence of internal facial features in the visual periphery, but not the orientation of these features. This confirms that observers are sensitive to the *presence* but not the *organisation* of facial features when visual acuity is limited in this way. In contrast to Experiment 11, these findings were now observed under conditions in which stimulus display time was limited to just 200 milliseconds, to reduce the possibility of covert attention shifts to these targets. These findings suggest that *some* information about internal features is available during the localisation of faces (i.e., their presence), but details such as the orientation of these features or their configuration of two eyes above a nose and mouth are not. This indicates that the detection process does not depend critically on this information. Instead, the findings of the experiments reported here point to a detection template that is based on general face-shape information. This is discussed in the next section.

General Discussion

This chapter introduced a novel gaze-contingent paradigm to dissociate the processes of face detection in the periphery and at fixation over a series of four experiments. Detection of faces that changed identity was comparable to faces that did not change, yet performance on response time and accuracy declined when faces disappeared between peripheral viewing and central fixation (Experiment 9). Experiment 10 revealed that response accuracy and time to first fixation performance increased for faces in which internal features were initially removed, but were added at fixation, but such performance declined when faces were initially presented intact, with internal features removed at fixation. Furthermore, featureless faces were distinguished faster and more accurately from intact faces when presented at opposing sides of peripheral vision, compared to faces where internal features were rotated 90º (Experiments 11 & 12).

Several insights emerge from these experiments. Firstly, it was revealed that detection is not sensitive to information such as a person's identity, but faces disappearing in the periphery causes a decline in performance due to the face not being present at the decision-making stage of detection (Experiment 9). Furthermore, internal features are not utilised in the initial localisation of faces as faces were fixated quickly irrespective of the initial presence of internal features, but internal features are likely to play a role in classification as accuracy decreased once those features were removed (Experiment 10). These findings were furthered in Experiments 11 and 12, in which observers do not seem to be sensitive to the *organisation* of internal features in the periphery, but are sensitive to their *presence.* This indicates that although 'face-like' features (e.g., the presence of eyes, nose and mouth, but not in a typical configuration) can be *seen* in the periphery, they are not utilised in detection. These findings converge with the evidence from Chapter 3, in that faces that preserve a basic colour-shape template can be detected with equal efficiency, regardless of the organisation of internal features (Pongakkasira & Bindemann, 2015).

The current experiments expand on the previous research with the use of eye-tracking to dissociate the features necessary *during* the detection process. The previous literature identified that faces are detected in peripheral vision (Boucart et al., 2016; Crouzet, Kirchner & Thorpe, 2010; Little, Jenkins & Susilo, 2021), and also concluded that manipulations of facial characteristics (such as shape, colour, features configuration) can hinder this process (Lewis & Edmonds, 2003; Pongakkasira & Bindemann, 2015; Bindemann & Lewis, 2013; Bindemann & Burton, 2009). What was less clear, however, was the extent to which such features were *perceived*, and which were *used* in the periphery to facilitate detection. Experiment 10 determined that faces with internal features disappearing in peripheral vision are detected less efficiently than intact faces, yet initially featureless faces that gain features are detected equally, if not more than, intact faces with no changes. This indicates that two differing processes may be involved in face detection that rely on different information. Firstly, such internal features seem not to be necessary in order to be drawn towards a face stimulus, as initially featureless faces were detected with equal efficiency to no-change faces. This is consistent with other research, in that such features may not be perceived in the periphery due to limitations in visual acuity (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005; Brown, Huey, & Findlay, 1997). Secondly, a decision-making process may occur at fixation in order to decide if the stimulus presented matches a cognitive template of a human face. It is during this process that internal features are utilised, which can be perceived clearly in foveal vision. This is also reflected in brain activity data in other studies, in which stronger responses to centrally presented faces are attributed to perceived facial detail (Levey et al., 2001).

This finding was furthered by Experiments 11 and 12, which demonstrated that peripheral vision is sensitive to the presence of internal features, but not their organisation. Faces with the internal features rotated were substantially more challenging to discriminate against intact faces, especially at greater eccentricities.

In contrast, featureless faces were distinguished much more efficiently, although performance declined at greater distances. In combination with the results of Experiment 10 (and the experiments in Chapter 3), this indicates that although internal features can be vaguely perceived at greater eccentricities, they are not utilised in the initial detection process. Such features are utilised, however, in a secondary decision-making process once processed in the central view. Here, face-like configurations, of two eyes, a mouth and a nose, are prioritised to classify face stimuli, as this is likely to be constituted as a face. This accounts for the decreased accuracy in identifying internal-disappear faces during the decision-making stage once faces were fixated (Experiment 10).

What is harder to account for in Experiment 10, however, is the enhanced detection of faces where the internal features were initially absent and then appeared. Here, two possible explanations are suggested, appertaining to detection theory and the methodology used. The first possibility is that internal features might actually *hinder* the detection process. That is, a basic, skin-coloured template (e.g., a consistent height-to-width ratio, with natural skin tones) may produce a more powerful initial draw than full featural faces (Bindemann & Burton, 2009; Pongakkasira & Bindemann, 2015). As internal features are not a precise face cue in the periphery, it is conceivable that a simpler template is employed so as to not be misled by any ‘face-like’ stimuli. This theorising draws in part on the phenomenon of face *pareidolia*, whereby non-face objects are sometimes perceived as faces (Omer, Sapir, Hatuka, & Yovel, 2019; Takahashi & Watanabe, 2015; Keys, Taubert, & Wardle, 2021). It is interesting to note that these pareidolia typically reflect the opposite stimulus characteristics that appear to be important in the current experiments, by capturing internal “features” such as two high contrast dots and a horizontal line that might represent a pair of eyes and a mouth (Omer et al., 2019), whereas a typical face outline and skin-colour tones are usually absent. In addition, experiments on the perception of pareidolia typically present these stimuli at or near fixation (Omer et al., 2019; Takahashi & Watanabe, 2015), and when the task requires search for a target in visual arrays, human faces outperformed these face-like objects (Keys et al., 2021). Therefore, if our visual system were to be drawn by ‘face-like’ pareidolic stimuli, this could potentially be distracting away from the visual information most useful to detect *actual* faces. Internal features may therefore be useful for detection in the absence of other facial information but are not essential in typical face detection. Additionally, internal features may not act as ideal detection candidates due to such variability in appearance. Individual identity means features differ in size and shape, which can be utilised in other face processes, such as recognition, but this may not be ideal for detection (Ellis, Shepherd, & Davies, 1979; Jarudi & Sinha, 2003). Furthermore, as such features are less visible in the periphery, they are subject to greater interference from irregularity in luminance (e.g., Lewis & Edmonds, 2003), and face view (i.e., features less, or not visible at all) (e.g. Bindemann & Lewis, 2013), and therefore convey a less standardised template compared to basic shape and skin colour.

Internal features are also potentially subject to skew or distortion in our periphery. An example of this, which also has the ability to impact the current experiment, is the flash distortion effect. When faces are flashed consistently and rapidly in the peripheral vision, this can cause facial features to appear distorted, such as features becoming blurred or disproportionate (Balas & Pearson, 2019; Bowden, Whitaker, & Dunn, 2019; Tangen, Murph, & Thompson, 2011). This distortion is also stronger at greater eccentricities in the periphery (Balas & Pearson, 2019; Bowden, Whitaker, & Dunn, 2019). Consequently, this effect may contribute to the results found at 10^{o} of eccentricity in both Experiments 11 & 12, as this effect is found from 6^{o}-8^{o} (Balas & Pearson, 2019). The current research does reduce this possibility by presenting drift check displays in between each trial (Tangen, Murph, & Thompson, 2011). Additionally, the ideal presentation for this effect is 4-5 faces per second, which was not achievable with the interruption of drift checks (Tangen, Murph, & Thompson, 2011). Therefore, although it is acknowledged that this effect could potentially influence the difficulty in distinguishing faces in Experiments 11 and 12, these effects should be minimised. Though, as such distortion can occur with the viewing of faces in the periphery, this would reduce their effectiveness as part of a cognitive template for detection. This distortion of features, the hindrance of features in the current research, and evidence that ‘face-like’ internal feature stimuli are detected less efficiently than human faces, collectively suggest that internal features may not be the most efficient candidates for face detection processes. Consequently, it is possible that a colour-shape template, based on circular skin-coloured tones but *without* internal features, might be a better candidate for finding human faces in the environment.

Alternatively, it is possible that the detection advantage of faces, in which the internal features were initially absent, could be attributed to the experimental design. As reported in Experiment 9, participants spontaneously claim to have seen ‘flashing’ in the stimuli during the experiment. This effect could be due to perceiving the change in the periphery, even if they are not aware that the change has occurred. This could account for changing faces and internal-appeared faces being detected more accurately, as this perceived flash draws attention towards such stimuli (Posner & Cohen, 1984; Laubrock, Engbert & Kliegl, 2005; Muller &

Rabbitt, 1989). However, this was not found with the removal of features, that would also supposedly create a flash effect. Either this strengthens the case that internal features play a crucial role in the decision-making process that overrides this initial draw, or the alternative explanation of features hindering detection holds.

The differentiation of detection processes goes some way in explaining the inconsistent findings from previous studies in the field, with the potential to clarify some of the debate over the contributions of internal and external features in face detection. Manipulating facial features, such as blurring, obscuring, reconfiguring internal features, or inverting and rotating external features, has given insight into the potential contributions of internal and external features to detection (e.g., Hershler & Hochstein, 2005; Lewis & Edmonds, 2003). All aforementioned manipulations decrease detection performance, so there is a lack of clarity on which, if any, features are essential for such processes (such effects are discussed and compared further in Chapter 5). Chapter 3 revealed that internal configuration does not contribute to detection performance compared to external configuration, but this did not account for why previous manipulations, such as scrambling features, were still found to hinder detection. The current research provides a clearer explanation for this effect, by suggesting that (at least) two distinct processes, that use different templates, occur during detection. Therefore, unless both templates are fulfilled, i.e., basic colour-shape for initial detection, and a feature configuration for classification, detection will deteriorate at one step of this process. This starts to build a coherent theory that accounts for diminished performance rates when either height-to-width ratios and colour are manipulated (Bindemann & Burton, 2009; Pongakkasira & Bindemann, 2015), when internal features are manipulated (Lewis & Edmonds, 2003), and why the most rapid rates occur when both are intact

(Hershler & Hochstein, 2005). These studies are discussed and compared further in Chapter 5.

The contribution of display background on the discrimination of faces was also considered in Experiments 11 and 12. Both experiments were consistent in their findings that faces are discriminated more efficiently on blank displays. This is consistent with Chapter 3, which concluded that detection was easier with decreased display complexity. Of significant note, however, is that featureless faces were distinguished more accurately on scene displays than blank displays. This is further evidence that a skin-colour template may be more successful for finding faces in complex displays that represent natural environments. This simple face template seemingly stands out from the detailed display, making it more salient to observers. As this was not found with internal-rotated features, this suggests that such facial detail may be limiting the draw towards the face to an extent, which is consistent with the findings from Experiment 10. This may also account for some of the inconsistencies in face detection research when presenting different faces in different displays, such as the influence of inverting faces, or presenting them in profile view, on blank displays or in scenes (e.g., Bindemann & Lewis, 2013) (as explored in Chapter 3).

The current findings start to differentiate the process involved *during* the detection of faces, and the facial features that facilitate these processes. Findings converge with previous research, suggesting that a basic colour-shape template is used for the initial detection of a face (Bindemann & Lewis, 2013; Chapter 3). This research furthers this by suggesting that once a face is located, a feature-specific template is used during the decision-making stage in order to classify faces. As visual acuity declines in the periphery, observers are not sensitive to the configuration of finer facial details such as internal features, and therefore these details are not utilised in peripheral detection. Yet, once faces can be viewed in the high-acuity regions of central vision, the additional detail of internal features can be utilised. Such findings emphasise the necessity to distinguish certain processes in research. The current studies start to define the boundaries for these processes, and how these may compare to other types of face perception.

Chapter 5:

Summary, Conclusions and Future Research

5.1 Summary and Conclusions

This thesis explored the features involved in the processes of face detection, and the contribution of different display contexts to such processes. Faces appear to hold a special status for the visual system whereby they are able to draw attention spontaneously, are preferentially viewed in natural scenes, and act as distractors during the search for other objects (Birmingham, Bischof & Kingstone, 2009; Bindemann, Scheepers, Ferguson & Burton, 2010; Langton, Law, Burton & Schweinberger, 2008; Crouzet, Kirchner & Thorpe, 2010). The use of the flicker paradigm, for example, demonstrates that changes to faces are detected more rapidly and accurately than other objects, indicating that attention is initially directed towards faces. Yet several task-dependent factors (e.g., the number of items displayed, the type of stimuli presented and the type of display) may also influence this attentional bias, bringing into question how this reflects real-world processing (see Palermo & Rhodes, 2003).

The face advantage extends to the detection of faces, in which faces are found faster than other objects and animals (Crouzet, Kirchner & Thorpe, 2010; Hershler & Hochstein, 2005; Di Giorgio, Turati, Altoè, & Simion, 2012; Maylott, Sansone, Jakobsen, & Simpson, 2021; Simpson, Maylott, Leonard, Lazo, & Jakobsen, 2019; Yang, Shih, Cheng, & Yeh, 2009). Faces can also be detected at greater eccentricities in the periphery than other objects (Rieth et al., 2011; Crouzet, Kirchner & Thorpe, 2010; De Lissa, McArthur, Hawelka, Palermo & Mahajan, 2019; Boucart et al., 2016). Therefore, faces seem to hold priority in the visual system, but what is driving this advantage is unclear. Detection seems to be optimal when faces are presented in a frontal view, upright orientation, and in full colour (Bindemann & Lewis, 2013; Pongakkasira & Bindemann, 2015; Bindemann &

Burton, 2009; Lewis & Edmonds, 2003). Once this is disrupted - for example, through inversion, the obscuring of features, or changes in face view - detection performance decreases (Bindemann & Lewis, 2013; Lewis & Edmonds, 2003; Hershler & Hochstein, 2005). However, there is a lack of consensus on which features are driving this detection process, as mixed evidence suggests that both the internal (e.g., eyes, mouth, nose) and external (e.g., shape, colour, hair) features provide separate contributions (Hershler & Hochstein, 2005).

One potential reason for this mixed evidence reflects methodological differences in the mode of stimulus presentation between different studies. The displays in which faces are presented are starting to be identified as a possible driver of such disparities. For example, the detection of frontal and profile faces appears to be comparable in blank displays, yet a discrepancy in the detection of these stimuli emerges once these are presented in more complex scenes (Bindemann & Lewis, 2013). Although this indicates that display impacts the conclusions we draw, there is a lack of systematic research into how display type may influence the process of face detection.

This thesis investigated this over a series of 12 experiments. Firstly, Chapter 2 examined the attentional draw of people in naturalistic scenes, and how the frequency of their appearance influences attentional bias. Across three experiments, a flicker paradigm was used to determine how person presence draws attention away from the detection of changes that occur elsewhere in visual displays. Participants were instructed to indicate whether a change was occurring or not within a scene, where a person was either present or absent. Naturalistic photos were used, containing full-bodied people, to avoid any semantic ‘out of place’ effects and give more insight into real-world processing. As changes were only occurring to non-face items in the scene, the task purposefully directs attention away from the person, to see if attention would still be drawn even when task-irrelevant. If change-detection performance declined in person-present scenes, then this would indicate that the person in the scene was drawing attention spontaneously in the task. Furthermore, the frequency with which people appeared in these scenes was manipulated to provide insight into how such task-dependent factors impact attentional biases. Thus, observers were asked to detect a change in images that were equally likely to contain a person or not (Experiments 1 & 2). Then, the frequency with which people appeared was manipulated to either 25%, 50% or 75% of trials, to determine whether person prevalence affected their relevance to the task (Experiment 3).

The experiments consistently showed that people in naturalistic scenes draw attention spontaneously. In Experiment 1, accuracy performance for detecting changes declined when a person was present in the scene compared to when a person was absent. Although the overall accuracy rates for Experiment 1 were relatively low (56% accuracy overall), Experiment 2 replicated the results with an easier version of the task with improved levels of performance (66% accuracy). Therefore, it is inferred that attention must be drawn towards the person in the scene, distracting from finding the changes occurring in other areas of the display. This is consistent with previous research, which has shown that changes to faces are detected faster than other objects (e.g., Ro, Lavie & Russell, 2001; Palermo & Rhoeds, 2003). Furthermore, in Chapter 2 this effect was dependent on person frequency, as it diminished as person prevalence increased. Specifically, Experiment 3 showed that attention was only distracted when a person appeared for either 25% or 50% of the experiment, but not when occurring 75%. This is in contrast to previous research suggesting that more frequent targets are better detected, whereas infrequent targets are often missed (Biggs, Adamo & Mitroff, 2014; Beanland, Le, & Byrne, 2016; Wolf et al., 2007; Goodwin et al., 2015; Menneer et al., 2010). This speaks to the importance of social stimuli in our visual environment, as such stimuli can capture attention at low frequency. Furthermore, this highlights how task demands can influence attentional biases. It is suggested that the person present becomes noticeably task-irrelevant in the high-frequency condition, as they *more clearly* do not contribute to finding the change, and therefore can be actively ignored. This reinforces the notion that faces and people have the ability to draw attention *spontaneously* but not draw attention *automatically*, and additionally highlights the importance of task demands in attentional research (Bindemann et al., 2007).

Chapter 2 established that faces and people draw attention spontaneously in naturalistic scenes, as well as highlighting the importance of task demands on research findings. Yet, these experiments cannot give insight into what is driving this advantage. Previous research indicates that person detection is driven strongly by the face (Bindemann et al., 2010; Langton, Law, Burton & Schweinberger, 2008; Ro, Russell & Lavie, 2001). Therefore, Chapter 3 examined the facial features that drive detection, and how display context influences the ability to detect faces. In a series of experiments, the detection of frontal faces was compared with profile faces (Experiment 4), faces rotated by 90° (Experiment 5) and inverted faces (Experiment 6), when these were displayed on either blank backgrounds, within an array of other objects, or embedded within scenes, as such displays may reveal differences in detection processes (Bindemann & Lewis, 2013; Hershler, Golan, Bentin & Hoshstein, 2010). The contributions of internal and external features on detection were also investigated by either rotating the internal or external features 90º to create hybrid faces (Experiment 7). The influence of display complexity was also investigated further by creating Voronoi scenes, in which areas of a display are converted into smaller sections that summarise colour information across the image, to determine how visual clutter affected detection (Experiment 8). In these experiments, faces were either present or absent in the visual displays, and observers reported whether they saw a face in the display or not.

These experiments consistently demonstrate that the complexity of the display influences how faces are detected. Experiments 4, 5, 6 and 7 all confirmed that the manipulated faces (i.e., profile, inverted, and rotated faces) were detected slower and less accurately than frontal, upright faces when presented in arrays and scenes, whereas performance was comparable for all face types when these were presented on blank displays. This is in line with previous research which also demonstrated that scenes reveal detection differences between frontal and profile faces, which did not occur in blank displays (Bindemann & Lewis, 2013). Display complexity, therefore, alters the type of processes involved in detection, such as the amount of search needed and the parsing of faces from the visual background. This was furthered with Experiment 8, which employed Voronoi scenes and demonstrated that detection performance declined with an increase in scene complexity, demonstrating that more detailed scenes provide a better basis to dissociate differences in the detection of varying face stimuli. Additionally, the experiments revealed key facial characteristics that are driving face detection. Experiment 4, for example, showed that profile faces are detected slower and less accurately than frontal faces in scenes, but detection is comparable in blank displays, consistent with previous research (Bindemann & Lewis, 2013). Yet profile faces are visually different from frontal faces in a number of aspects, such as their overall shape, the visibility of features (e.g., one eye versus a pair of eyes), and the amount of skin and hair that is on display. Due to these stark differences between these face conditions, Experiment 4 does not reveal much about the *specific* features that are driving the differences in detection. Therefore, Experiments 5 and 6 presented frontal faces rotated by 90º and 180º respectively, in which the visual information of frontal faces remains the same, but only the orientation changes. Both experiments demonstrated declined performance in detecting the manipulated faces in scenes and arrays. Rotated faces disrupt the face shape, indicating that this would hold importance in detection, however, as inverted faces generally hold a similar external shape to upright faces, this could imply alternatively that internal feature configuration is a significant driver (Brown, Huey, & Findlay, 1997; Lewis & Edmonds, 2003; Lewis & Edmonds, 2005). To examine this directly, Experiment 7 selectively manipulated internal and external features, by rotating one or the other by 90º, to determine whether general face shape or internal featural configuration holds greater importance in detection. Here, faces with the internal features rotated were detected at the same rate as normal upright faces, yet when the external features were rotated, performance declined. This suggests that the configuration of internal features is not driving face detection, but rather the external features. This is in contrast to previous research which indicated that internal features hold importance (Lewis & Edmonds, 2003), or that internal and external features contribute to detection equally (Hershler & Hochstein, 2005). The current research, however, may convey a more convincing argument for the importance of external features in detection, due to the separate manipulation of features, without removing them entirely (Hershler & Hochstein, 2005).

Both Chapters 2 and 3 demonstrated an advantage of being drawn towards person stimuli in naturalistic scenes. The use of scenes in Chapter 2 aimed to reveal whether person stimuli can draw attention spontaneously in such displays. This is in contrast to previous research that commonly used array designs to consider the attentional draw to face stimuli (Palermo & Rhoeds, 2003; Ro, Russell, & Lavie, 2001). The experiments in Chapter 2 showed that scene displays are sufficient to elicit the attentional draw of person stimuli. The importance of display type was revealed in Chapter 3, in which the pattern of detection was altered with blank displays compared to arrays and scenes. These experiments not only demonstrated the importance of display context when presenting faces but also showed that such contexts reveal key characteristics involved in detection processes. Collectively, both chapters illustrate the importance of using scene displays in attention and detection research, as complex visual displays are necessary to elicit the differences in search that is necessary to provide a fuller understanding of such processes. Yet, the methods used in the previous chapters provide indirect measures that could not reveal the locus at which specific visual information, such as internal or external features, becomes important for face detection. Therefore, the extent to which facial details are utilised in the processes of detection remained unclear.

With this in mind, the final experimental chapter examined how internal and external features drive detection during the localisation of faces in the visual periphery. A gaze-contingent eye-tracking paradigm was used, whereby changes could be made to faces in scenes as participants saccaded towards them, allowing for the separation of localisation and fixation of faces. How visual changes are processed in the periphery was first investigated by either having faces stay the same, disappear or change in identity as observers saccaded towards them

(Experiment 9). If changes in the periphery are not noticed, then performance at the point of the detection decision should have declined when faces are removed, but changes in identity should not influence detection performance. On the other hand, if such changes were perceived in the periphery, influencing detection performance, then this would have impacted the interpretation of results in the proceeding experiment, as the changes themselves could be driving the differences in detection between faces. The following experiment manipulated the appearance and disappearance of internal features as saccades were made towards the faces (Experiment 10). This was to determine whether internal features were utilised in the localisation of faces. A decline in performance when internal features are not visible would signify their importance in the detection process, whereas consistent performance regardless of internal feature presence would suggest they are not utilised during detection. Finally, the extent to which internal features could be perceived in the periphery was investigated by presenting both normal and manipulated faces (either featureless with no internal features; or with internal features rotated 90º) in the periphery and asking observers to distinguish between them by identifying the normal face (Experiments 11 & 12). Decreased performance for internally rotated faces compared to blank faces would indicate that although internal features are seen peripherally, the visual system may not be sensitive to featural orientation. This gives insight into the extent features are *seen* compared to how they are *used* in detection.

The experiments revealed a distinction between features important for the localisation of faces, and those important for their classification (i.e., deciding that a face is indeed a face). Experiment 9 indicated that face presence is necessary at the decision-making stage, as the disappearance of a face reduced detection performance. Yet identity changes did not have this effect, suggesting that peripheral detection is not sensitive to information such as person identity. Experiment 10 revealed faces where internal features were initially removed, were detected equally, and perhaps enhanced detection performance, than intact faces with no change, and faces where the internal features disappeared at fixation. This not only suggests that internal features are not used in the localisation of faces, but such features might even hinder the detection of faces at this stage. This may be an advantage so that the cognitive system for face detection is not distracted by false positives, such as ‘face-like’ object stimuli (Omer, Sapir, Hatuka, & Yovel, 2019; Takahashi & Watanabe, 2015; Keys, Taubert, & Wardle, 2021), or due to distortions that skew the perception of internal features when rapidly viewing faces in the periphery (Balas & Pearson, 2019; Bowden, Whitaker, & Dunn, 2019; Tangen, Murph, & Thompson, 2011) or lack of visual acuity in the periphery (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005). Furthermore, when internal features were removed at fixation, performance decreased. This indicates that although internal features are not utilised in the localisation of faces, the presence of eyes, nose and mouth are vital for face classification when deciding whether a face, is indeed, a face. This explains some disparity in previous findings on the importance of internal and external features, as research was not able to distinguish the process of localisation and classification (Lewis & Edmonds, 2003; Hershler & Hochstein, 2005). These results were consolidated by Experiments 11 and 12. In these experiments, observers were sensitive to the presence of internal facial features but not their orientation, with featureless faces being distinguished better than faces in which the internal features were rotated. This indicates that internal features are unlikely to contribute to initial face detection, perhaps because the lack of visual acuity in the periphery makes their visibility limited (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005).

These findings clarify the results of Experiment 7, which showed that faces in which the internal features are rotated are detected equally as fast as normal frontal faces. Yet, Experiment 7 did not show a classification cost with internally rotated faces that would be expected at the decision-making stage. Experiments 11 and 12 may provide a further explanation for this, as the finer details of facial features are difficult to perceive at greater eccentricities (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005). Therefore, it is suggested that observers make quick decisions based on peripheral vision in Experiment 7. Experiment 10 also revealed that faces were fixated on less than half of trials (47%) but performance was still high (96%) indicating that responses were occurring before fixation. This therefore might explain how the internally rotated faces in Experiment 7 were mistaken for full faces, as if they were classified before fixation, the rotated internal features could be mistaken for full faces. This effect may not have been observed in Experiment 10, in which faces were presented without internal features, rather than having rotated visible features in view. This also builds a stronger case for internal features being utilised in the classification of faces, rather than their detection.

Chapter 4 successfully builds on Chapter 3, by clarifying the role of internal features in face detection whilst separating the processes of localisation and classification. In Chapter 3, Experiment 7 demonstrated that the configuration of internal features is not essential for face detection. Experiment 10 reinforced this, as faces with no changes and faces with internal features initially (then disappearing) were fixated at a similar rate, during the localisation stage. However, accuracy decreased when the internal features were removed, during classification. This indicates that although internal features are not necessary for the initial localisation of the face, they gained importance in the decision-making stage. As previously mentioned, internal feature configuration is hard to distinguish in peripheral vision, yet such features are still necessary for faces to be perceived as ‘face-like’ closer to fixation. This identifies the separate processes of detection that should be considered in such research.

The suggestion of differing processes within detection indicates how external and internal features contribute separately to the localisation of the face and its classification during the detection process. A more simple, colour shape template seems to be utilised for the localisation of a face, whereas a more detailed template with a typical internal featural configuration is necessary to then confirm that a located face candidate is in fact a face. This shows consistency with previous research which indicates that both skin-coloured tones and normal height-to-width ratios of faces hold importance in detection (Bindemann & Burton, 2009; Pongakkasira & Bindemann, 2015). However, faces without such a colour shape template, such as schematic faces, or face pareidolia, are still found to be located faster than non-face items (Omer, Sapir, Hatuka, & Yovel, 2019; Keys, Taubert, & Wardle, 2021; Takahashi & Watanabe, 2015). This also suggests that when the features of the primary colour-shape template are not met, a secondary template consisting of the other identifying features can be used. As faces of this nature are still detected less efficiently than full-coloured photographs of faces, this indicates that the colour-shape template holds priority (Keys, Taubert, & Wardle, 2021; Takahashi & Watanabe, 2015).

The findings also contribute to the previously inconsistent debate over whether internal features or external features are necessary for face detection (e.g., Hershler & Hochstein, 2005; Lewis & Edmonds, 2003). Previous evidence was mixed, with different studies highlighting the importance of internal and external features independently, suggesting that they may equally contribute to the detection process (Hershler & Hochstein, 2005). Here, however, it is clear that the previous research mentioned was not able to separate the processes of localisation and classification, which explains why it seems these features have similar contributions. Not only does this give insight into this debate, but also how the methods used in such research may cause inconsistencies in findings. Research in which the face encompasses a large proportion of the display may lack the search element required for detection, and therefore elicit similar effects to blank displays, as found in Chapter 3 (Lewis & Edmonds, 2003). The way internal and external features are presented or manipulated also influences findings. For example, if external features are 'removed' by blurring the edges of faces, then the basic shape may remain intact, and this additionally does not account for the role of colour (Hershler & Hochstein, 2005). Furthermore, the removal of features entirely enhances detection but hinders classification (Hershler & Hochstein, 2005). Such confounding variables create difficulties in distinguishing which features contribute to detection processes. The current experiments circumvented this issue by manipulating the appearance of features at these different stages, meaning such processes could be separated.

Not only does this theory indicate that internal features are not utilised in peripheral vision, but suggests that the presence of such features might potentially hinder detection. Faces in which the internal features were initially removed during the localisation phase elicited greater detection performance than when internal features were removed near fixation and when faces remained intact. This indicates that the colour shape-template without features has more draw, due to standing out more clearly in scene displays (See Experiments 11 & 12). This theory holds logic when considering other phenomena such as face pareidolia in which the perception of internal features would potentially distract attention too often to false targets (Omer, Sapir, Hatuka, & Yovel, 2019; Takahashi & Watanabe, 2015; Keys, Taubert, & Wardle, 2021), and flash distortion effects in which features may be perceived as skewed in the periphery (Balas & Pearson, 2019; Bowden, Whitaker, & Dunn, 2019; Tangen, Murph, & Thompson, 2011) and are therefore unreliable (See Chapter 4 General Discussion). Overall the evidence converges on a theory that internal features are not useful in the initial detection of faces, which should be considered in the evaluation of future research.

The findings of this thesis have clear implications for how we use methodologies in distinguishing different cognitive processes in face detection. The processes of detection were explored to reveal the features necessary for face detection, whilst determining the display contexts that facilitate this. Demonstrating that presenting faces on blank displays elicits different findings than when presented in scenes indicates that *how* detection is investigated determines *what* is learned about the detection process. Therefore, more complex displays are best to facilitate the search that is more representative of real-world circumstances. The use of such scene displays revealed that detection relies more upon external facial features, such as colour and shape, rather than being sensitive to internal features. This was explored further by separating the processes involved during the detection of faces. During the localisation of faces, a basic head shape and skin colour (and without internal features) is sufficient to detect a face in the periphery. Yet during the classification of faces, the presence of intact internal features (eyes, nose and mouth in a typical configuration) is utilised. When considering the lack of visual acuity in the periphery, it becomes apparent why internal features may not be of importance in the detection processes, as they cannot be clearly perceived (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005; Westheimer, 1965; Kalloniatis & Luu, 2007).

These findings point to a theory in which two distinct cognitive templates are at play in the detection process. Firstly, a colour shape template is used to localise a face in the periphery, which is more visible and easier to detect. This converges with other research that indicates distortions in the colour or shape of faces impede performance (Bindemann & Burton, 2009; Pongakkasira & Bindemann, 2015). Secondly, a more feature-specific template arises when classifying faces, which relies on a typical configuration of facial features. This occurs closer to fixation when such features are more visible. These templates demonstrate that the processes of detection are distinct from other forms of face processing, such as identity and emotion recognition that do primarily rely on internal features (see Ellis, Shepherd, & Davies, 1979; Toseeb, Keeble, & Bryant, 2012; Wegrzyn, Vogt, Kireclioglu, Schneider, & Kissler, 2017). Thus, detection should be researched and understood in its own right, with a focus on the use of naturalistic scenes, whilst accounting for the potential different processes occurring during detection.

5.2 Future Directions

The current research raises further enquiries to be pursued. Chapter 2 demonstrated that person stimuli have the ability to draw attention spontaneously. However, based on the design of the experiments in this chapter, it is not possible to specify *when* this attention draw occurs. For example, whether this attentional draw is automatic upon the first appearance of the stimuli or does such stimuli repeatedly draw attention during the search for the change. Experiment 3 may give some insight into this, as it suggests attention can be shifted to favour the task demands (also see Bindemann et al, 2007), and therefore person stimuli are not drawing *automatic* attention. Yet, alternative methods, such as eye-tracking, may give more insight into where and how much attention is being directed during the task, as attention and eye movements have been closely associated (Findlay, 2005; Hoffman & Subramaniam, 1995; Peterson, Kramer & Irwin, 2004).

Chapter 3 has clear implications for how face detection is researched and how conclusions are drawn from such research. Disparities between the detection of frontal faces and manipulated faces were only found to occur in array displays and scenes (i.e., complex displays). This has implications when considering forms of research that utilise blank displays to avoid noise in the data. For example, research using methods such as EEG or fMRI may not be an optimal measure of face detection, and should therefore be cautious when describing such research as ‘detection’ (e.g. Gao, Xu, Zhang, Zhao, Harel, & Bentin, 2009; Looser, Guntupalli, & Wheatley, 2013; Moulson, Balas, Nelson, & Sinha, 2011). Having to parse faces from a more complex background may require a combination of information (such as colour and shape) that is not a template that needs to be utilised when viewing faces on blank displays. Research suggests that brain activity differs when presenting visual scenes compared to faces in isolation, suggesting that additional processes occur when viewing different displays that may impact overall processing (Epstein & Kanwisher, 1998). This difference was also found when comparing multi-item arrays and scenes (Epstein & Kanwisher, 1998). Therefore, the use of naturalistic scenes may provide a better representation of how we process the environment in the real world. However, it should also be noted that this still does not represent the dynamic nature of real environments, and therefore alternative methods, such as the use of videos, or mobile eye-tracking may also pose an important contribution (Foulsham, et al, 2010; Itti, 2005).

A similar critique to Chapter 2 can be applied to Chapter 3 in terms of eye-tracking creating a clearer depiction of the detection processes. However, this was addressed in Chapter 4. Chapter 4 provides a coherent contribution to the theory of face detection. Here it was discovered that internal features of a face are not utilised in the initial localisation of a face, suggesting that a more basic colour-shape template is necessary for detection. Furthermore, internal features are applied in a secondary process, in which faces must be classified in the decision-making stage. This was furthered by the finding that the human visual system is not necessarily sensitive to the organisation of internal features in the periphery, and therefore they would not necessarily be useful during detection. This informs how face detection research should be conducted and considered in the future. For example, when considering the internal versus external feature debate, it should be taken into account whether the methods used distinguish between the localisation of the faces and the classification. How faces are portrayed and their manipulations should also be considered in how this can inform research. Confounding variables, such as the size of the faces, how intact the colour-shape template is, and the visibility of features, alter these localisation and classification stages of detection. Therefore, this must be considered in experimental design and when drawing conclusions.

As these are preliminary findings on this theory, the concept of internal features hindering detection should be explored further to clarify *how* and under *what* circumstances this colour shape template holds its draw. For example, investigating whether oval-like shapes that are skin coloured elicit greater detection performance than items with the face pareidolia effect. Furthermore, it would be insightful to have a clearer separation of the processes of detection, e.g., the localisation and classification phases, and at what points they occur. Broadly, this research indicates that localisation is peripheral, whereas classification occurs at fixation, or at least close to fixation. As not all faces were fixated (47% of faces present were fixated in Experiment 10), but overall performance was still high (96% overall accuracy in Experiment 10), this indicates that classification occurs close to fixation, when internal features are clearly visible, but does not have to be directly at fixation. Research into visual acuity could give more insight into the point at which features become clearly visible (Westheimer, 1965; Kalloniatis & Luu, 2007). Previous research indicates that face targets are difficult to distinguish at 7º of eccentricity (Brown, Huey, & Findlay, 1997), yet Experiments 11 and 12 demonstrated that at 5º of the visual field, distinguishing internal features can be challenging, and this is more so at 10º. As the face-like stimuli in Experiments 11 and 12 were still being mistaken as intact faces, it would suggest that such stimuli would be classified as faces, as demonstrated in Experiment 7. This could suggest that face classification occurs outside of foveal vision and occurs at parafoveal vision. Yet the point at which this occurs is hard to determine, as factors such as sizing and scaling of stimuli need to be taken into account (Melmoth, Kukkonen, Makela & Rovamo, 2000; Rousselet, Husk, Bennett & Sekuler, 2005). Neuronal responses to faces diminish at 10º of eccentricity, but this effect does not occur when the stimuli size is accounted for. Therefore, the classification of such faces would be subject to this, so it is still unclear as to what point these internal features are utilised for categorisation.

The detection colour shape template can be applied to other face phenomena, such as the other-race effects in face processing. The other race effect (or own-race bias) occurs as an ingroup bias, in which recognition of faces is improved for a person's own race or ethnicity, compared to other ethnic groups (see Meissner & Brigham, 2001). This effect has also been found in detection, in which detection performance increased for ingroup faces compared to outgroup faces (Prunty, Jenkins, Qarooni & Bindemann, 2022). The colour-shape template described in this thesis may contribute to the understanding of such effects. For example, it would be logical that an individual’s skin-colour template for detection matches one's own race, particularly if living in an environment with one's own ethnicity where they would statistically more likely to be encountered. As this occurs in such an early processing phase, this also helps contribute to understanding how this template is applied at the later stages of processing, such as identity recognition. Research on the other-race effects does note that the bias occurs even when the colour hue is adjusted (e.g., to black and white images), suggesting that these effects may not be due to skin colour per se (Prunty, Jenkins, Qarooni & Bindemann, 2022). However, adjusting hue decreased overall detection performance, suggesting that colour is optimal for the detection template (Prunty, Jenkins, Qarooni & Bindemann, 2022; Bindemann & Burton, 2009). Therefore, it is suggested that the colour shape template is the primary detection template, but in the absence of these features, other features can be utilised in detection. For example, in the case of the other race effect, tone and contrast of other features may be considered and driving the bias once colour cannot be accounted for.

Overall, this thesis proposes a cognitive template for face detection that should be considered when investigating the processes of face perception. The localisation of faces in the visual field seems to be driven by a template consisting of a skin-toned oval and is potentially most optimal when the internal features are *not* present. However, the internal features are used as part of a template that subsequently classifies faces, which occurs closer to fixation. Separating these processes not only explains disparities in the internal/external feature debate but also contributes to the understanding of other aspects of face processing, such as other-race effects. In turn, this builds a more coherent theory of face detection processes.

References

Alais, D., Xu, Y., Wardle, S. G., & Taubert, J. (2021). A shared mechanism for facial expression in human faces and face pareidolia. *Proceedings of the Royal Society B, 288(1954)*, 20210966. doi: 10.1098/rspb.2021.0966.

Balas, B., & Pearson, H. (. (2019). The Flashed Face Distortion Effect Does Not Depend on Face-Specific Mechanisms. *Scientific Reports, 9(1)*, 1612. doi: 10.1038/s41598-018-37991-9.

Bayle, D. J., Schoendorff, B., Hénaff, M. A., & Krolak-Salmon, P. (2011). Emotional facial expression detection in the peripheral visual field. *PloS One, 6(6)*, e21584. doi: 10.1371/journal.pone.0021584.

Beanland, V., Le, R. K., & Byrne, J. E. (2016). Object–scene relationships vary the magnitude of target prevalence effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 42(6)*, 766-775. doi: 10.1037/xhp0000183.

Bernard, P., Content, J., Deltenre, P., & Colin, C. (2018). When the body becomes no more than the sum of its parts: the neural correlates of scrambled versus intact sexualized bodies. *NeuroReport, 29(1)*, 48-53. doi: 10.1097/WNR.0000000000000926.

Biggs, A. T., Adamo, S. H., & Mitroff, S. R. (2014). Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. *Acta Psychologica, 152*, 158-165. doi: 10.1016/j.actpsy.2014.08.005.

Bindemann, M., & Burton, A. M. (2008). Attention to upside-down faces: An exception to the inversion effect. *Vision Research, 48(25),* 2555-2561. doi: 10.1016/j.visres.2008.09.001.

Bindemann, M., & Burton, A. M. (2009). The role of color in human face detection. *Cognitive Science, 33(6),* 1144-1156. doi: 10.1111/j.1551-6709.2009.01035.x.

Bindemann, M., & Lewis, M. B. (2013). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin & Review, 20*, 1140-1145. doi: 10.3758/s13423-013-0445-9.

Bindemann, M., Burton, A. M., & Jenkins, R. (2005). Capacity limits for face processing. *Cognition, 98(2),* 177-197. doi: 10.1016/j.cognition.2004.11.004.

Bindemann, M., Burton, A. M., Langton, S. R., Schweinberger, S. R., & Doherty, M. J. (2007). The control of attention to faces. *Journal of Vision, 7(10),* 15-15. doi: 10.1167/7.10.15.

Bindemann, M., Scheepers, C., & Burton, A. M. (2009). Viewpoint and center of gravity affect eye movements to human faces. *Journal of Vision, 9(2),* 7-7. doi: 10.1167/9.2.7.

Bindemann, M., Scheepers, C., Ferguson, H. J., & Burton, A. M. (2010). Face, body, and center of gravity mediate person detection in natural scenes. *Journal of Experimental Psychology: Human Perception and Performance, 36(6),* 1477-1485. doi: 10.1037/a0019057.

Birmingham, E., Bischof, W. F., & Kingstonee, A. (2008). Gaze selection in complex social scenes. *Visual Cognition, 16(2-3),* 341-355. doi: 10.1080/13506280701434532.

Birmingham, E., Bischof, W. F., & Kingstonee, A. (2009). Saliency does not account for fixations to eyes within social scenes. *Vision Research, 49(24),* 2992-3000. doi: 10.1016/j.visres.2009.09.014.

Boucart, M., Lenoble, Q., Quettelart, J., Szaffarczyk, S., Despretz, P., & Thorpe, S. J. (2016). Finding faces, animals, and vehicles in far peripheral vision. *Journal of Vision, 16(2),* 10-10. doi: 10.1167/16.2.10.

Bowden, J., Whitaker, D., & Dunn, M. J. (2019). The role of peripheral vision in the flashed face distortion effect. *Perception, 48(1),* 93-101. doi: 10.1177/0301006618817419.

Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception, 26(12),* 1555-1570. doi: 10.1068/p261555.

Burton, A. M., & Bindemann, M. (2009). The role of view in human face detection. *Vision Research, 49(15),* 2026-2036. doi: 10.1016/j.visres.2009.05.012.

Burton, A. M., Bindemann, M., Langton, S. R., Schweinberger, S. R., & Jenkins, R. (2009). Gaze perception requires focused attention: Evidence from an interference task. *Journal of Experimental Psychology: Human Perception and Performance, 35(1),* 108–118. doi: 10.1037/0096-1523.35.1.108.

Calvo, M. G., & Nummenmaa, L. (2008). Detection of emotional faces: salient physical features guide effective visual search. *Journal of Experimental Psychology: General, 137(3)*, 471-494. doi: 10.1037/a0012771.

Calvo, M. G., Nummenmaa, L., & Avero, P. (2010). Recognition advantage of happy faces in extrafoveal vision: Featural and affective processing. *Visual Cognition, 18(9)*, 1274-1297. doi: 10.1080/13506285.2010.481867.

Cecchini, M., Aceto, P., Altavilla, D., Palumbo, L., & Lai, C. (2013). The role of the eyes in processing an intact face and its scrambled image: A dense array ERP and low-resolution electromagnetic tomography (sLORETA) study. *Social Neuroscience, 8(4)*, 314-325. doi: 10.1080/17470919.2013.797020.

Cerf, M., Harel, J., Einhäuser, W., & Koch, C. (2007). Predicting human gaze using low-level saliency combined with face detection. *Advances in Neural Information Processing Systems, 20.*

Chai, X. J., Ofen, N., Jacobs, L. F., & Gabrieli, J. D. (2010). Scene complexity: influence on perception, memory, and development in the medial temporal lobe. *Frontiers in Human Neuroscience, 4*, 1021. doi: 10.3389/fnhum.2010.00021.

Cooper, R. M., & Langton, S. R. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour Research and Therapy, 44(9)*, 1321-1329. doi: 10.1016/j.brat.2005.10.004.

Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. *Frontiers in Psychology, 2*, 342. doi: 10.3389/fpsyg.2011.00342.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. *Journal of Vision, 10(4)*, 16-16. doi: 10.1167/10.4.16.

Davenport, J. L., & Potter, M. C. (2004). Scene consistency in object and background perception. *Psychological Science, 15(8)*, 559-564. doi: 10.1111/j.0956-7976.2004.00719.x.

Davies, T. N., & Hoffman, D. D. (2002). Attention to faces: A change-blindness study. *Perception, 31(9)*, 123-1146. doi: 10.1068/p3391.

De Lissa, P., McArthur, G., Hawelka, S., Palermo, R., Mahajan, Y., Degno, F., & Hutzler, F. (2019). Peripheral preview abolishes N170 face-sensitivity at fixation: Using fixation-related potentials to investigate dynamic face processing. *Visual Cognition, 27(9-10)*, 740-759. doi: 10.1080/13506285.2019.1676855.

Di Giorgio, E., Turati, C., Altoè, G., & Simion, F. (2012). Face detection in complex visual displays: An eye-tracking study with 3-and 6-month-old infants and adults. *Journal of Experimental Child Psychology, 113(1)*, 66-77. doi: 10.1016/j.jecp.2012.04.012.

Dickins, D. S., & Lipp, O. V. (2014). Visual search for schematic emotional faces: Angry faces are more than crosses. *Cognition & Emotion, 28(1)*, 98-114. doi: 10.1080/02699931.2013.809331.

Donnelly, N., Humphreys, G. W., & Sawyer, J. (1994). Stimulus factors affecting the categorisation of faces and scrambled faces. *Acta Psychologica, 85(3)*, 219-234. doi: 10.1016/0001-6918(94)90036-1.

Downing, P. E., Bray, D., Rogers, J., & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition, 93(1)*, B27-B38. doi: 10.1016/j.cognition.2003.10.010.

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293(5539),* 2470-2473. doi: 10.1126/science.1063414.

Downing, P. E., Peelen, M. V., Wiggett, A. J., & Tew, B. D. (2006). The role of the extrastriate body area in action perception. *Social Neuroscience, 1(1),* 52-62. doi: 10.1080/17470910600668854.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96(3),* 433-458. doi: 10.1037/0033-295X.96.3.433.

Easterbrook, M. A., Kisilevsky, B. S., Hains, S. M., & Muir, D. W. (1999). Faceness or complexity: Evidence from newborn visual tracking of facelike stimuli. *Infant Behavior and Development, 22(1)*, 17-35. doi: 10.1016/S0163-6383(99)80003-4.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8(4)*, 431-439. doi: 10.1068/p080431.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature, 392(6676),* 598-601. doi: 10.1038/33402.

Farah, M. J., Wilson, K. D., Drain, H. M., & Tanaka, J. R. (1995). The inverted face inversion effect in prosopagnosia: Evidence for mandatory, face-specific perceptual mechanisms. *Vision Research, 35(14),* 2089-2093. doi: 10.1016/0042-6989(94)00273-O.

Findlay, J. M. (2005). Covert attention and saccadic eye movements. In *Neurobiology of Attention* (pp. 114-116). Academic Press. doi: 10.1016/B978-012375731-9/50024-0.

Fletcher-Watson, S., Findlay, J. M., Leekam, S. R., & Benson, V. (2008). Rapid detection of person information in a naturalistic scene. *Perception, 37(4),* 571-583. doi: 10.1068/p5705.

Foulsham, T., Cheng, J. T., Tracy, J. L., Henrich, J., & Kingstonee, A. (2010). Gaze allocation in a dynamic situation: Effects of social status and speaking. *Cognition, 117(3),* 319-331. doi: 10.1016/j.cognition.2010.09.003.

Gao, L., Xu, J., Zhang, B., Zhao, L., Harel, A., & Bentin, S. (2009). Aging effects on early-stage face perception: An ERP study. *Psychophysiology, 46(5),* 970-983. doi: 10.1111/j.1469-8986.2009.00853.x.

Gliga, T., Elsabbagh, M., Andravizou, A., & Johnson, M. (2009). Faces attract infants' attention in complex displays. *Infancy, 14(5),* 550-562. doi: 10.1080/15250000903144199.

Gluckman, M., & Johnson, S. P. (2013). Attentional capture by social stimuli in young infants. *Frontiers in Psychology, 4,* 527. doi: 10.3389/fpsyg.2013.00527.

Godwin, H. J., Menneer, T., Cave, K. R., Thaibsyah, M., & Donnelly, N. (2015). The effects of increasing target prevalence on information processing during visual search. *Psychonomic Bulletin & Review, 22*, 469-475. doi: 10.3758/s13423-014-0686-2.

Goodrich, R. I., & Yonelinas, A. P. (2020). The effects of face inversion on perceiving-and sensing-based change detection. *Journal of Experimental Psychology: General, 149(1)*, 79-93. doi: 10.1037/xge0000618.

Groen, I. I., Jahfari, S., Seijdel, N., Ghebreab, S., Lamme, V. A., & Scholte, H. S. (2018). Scene complexity modulates degree of feedback activity during object detection in natural scenes. *PLoS Computational Biology, 14(12)*, e1006690. doi: 10.1371/journal.pcbi.1006690.

Hadjikhani, N., Kveraga, K., Naik, P., & Ahlfors, S. P. (2009). Early (N170) activation of face-specific cortex by face-like objects. *Neuroreport, 20(4)*, 403-407. doi: 10.1097/WNR.0b013e328325a8e1.

Hershler, O., & Hochstein, S. (2005). At first sight: A high-level pop out effect for faces. *Vision Research, 45(13)*, 1707-1724. doi: 10.1016/j.visres.2004.12.021.

Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *Journal of Vision, 10(10)*, 21-21. doi: 10.1167/10.10.21.

Hietanen, J. K., & Leppänen, J. M. (2003). Does facial expression affect attention orienting by gaze direction cues? *Journal of Experimental Psychology: Human Perception and Performance, 29(6)*, 1228-1243. doi: 10.1037/0096-1523.29.6.1228.

Hietanen, J. K., & Nummenmaa, L. (2011). The naked truth: the face and body sensitive N170 response is enhanced for nude bodies. *PLoS One, 6(11),* e24408. doi: 10.1371/journal.pone.0024408.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics, 57(6),* 787-795. doi: doi.org/10.3758/BF03206794.

Hole, G. J., George, P. A., Eaves, K., & Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception, 31(10)*, 1221-1240. doi: doi.org/10.1068/p3252.

Hollingworth, A., & Henderson, J. M. (2000). Semantic informativeness mediates the detection of changes in natural scenes. *Visual Cognition, 7(1-3),* 213-235. doi: 10.1080/135062800394775.

Houtmans, M. J., & Sanders, A. F. (1983). Is information acquisition during large saccades possible? *Bulletin of the Psychonomic Society*, *21(2)*, 127-130. doi: 10.3758/BF03329974.

Itti, L. (2005). Quantifying the contribution of low-level saliency to human eye movements in dynamic scenes. *Visual Cognition, 12(6)*, 1093-1123. doi: 10.1080/13506280444000661.

Jarudi, I. N., & Sinha, P. (2003). Relative contributions of internal and external features to face recognition. *Artificial Intelligence Lab Publications*.

Jenkins, R., Lavie, N., & Driver, J. (2003). Ignoring famous faces: Category-specific dilution of distractor interference. *Perception & Psychophysics, 65(2)*, 298-309. doi: 10.3758/BF03194801.

Johnson, M. H., Dziurawiec, S., Ellis, H., & Morton, J. (1991). Newborns' preferential tracking of face-like stimuli and its subsequent decline. *Cognition, 40(1-2),* 1-19. doi: 10.1016/0010-0277(91)90045-6.

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object vision in a structured world. *Trends in Cognitive Sciences, 23(8),* 672-685. doi: 10.1016/j.tics.2019.04.013.

Kalloniatis, M., & Luu, C. (2007). Visual acuity. *Webvision: The Organization of the Retina and Visual System [Internet].*

Kanwisher, N. (2001). Faces and places: of central (and peripheral) interest. *Nature Neuroscience, 4(5),* 455-456. doi: 10.1038/87399.

Kelly, D. J., Duarte, S., Meary, D., Bindemann, M., & Pascalis, O. (2019). Infants rapidly detect human faces in complex naturalistic visual scenes. *Developmental Science, 22(6),* e12829. doi: 10.1111/desc.12829.

Kemp, R., Pike, G., White, P., & Musselman, A. (1996). Perception and recognition of normal and negative faces: the role of shape from shading and pigmentation cues. *Perception, 25(1),* 37-52. doi: 10.1068/p250037.

Kendall, L. N., Raffaelli, Q., Kingstonee, A., & Todd, R. M. (2016). Iconic faces are not real faces: enhanced emotion detection and altered neural processing as faces become more iconic. *Cognitive Research: Principles and Implications, 1(1),* 1-14. doi: 10.1186/s41235-016-0021-8.

Keys, R. T., Taubert, J., & Wardle, S. G. (2021). A visual search advantage for illusory faces in objects. *Attention, Perception, & Psychophysics, 83,* 1942-1953. doi: 10.3758/s13414-021-02267-4.

Kovács, G., Zimmer, M., Bankó, É., Harza, I., Antal, A., & Vidnyánszky, Z. (2006). Electrophysiological correlates of visual adaptation to faces and body parts in humans. *Cerebral Cortex, 16(5)*, 742-753. doi: 10.1093/cercor/bhj020.

Kreegipuu, K., Kuldkepp, N., Sibolt, O., Toom, M., Allik, J., & Näätänen, R. (2013). vMMN for schematic faces: automatic detection of change in emotional expression. *Frontiers in Human Neuroscience, 7,* 714. doi: 10.3389/fnhum.2013.00714.

Krombholz, A., Schaefer, F., & Boucsein, W. (2007). Modification of N170 by different emotional expression of schematic faces. *Biological Psychology, 76(3),* 156-162. doi: 10.1016/j.biopsycho.2007.07.004.

Kuehn, S. M., & Jolicoeur, P. (1994). Impact of quality of the image, orientation, and similarity of the stimuli on visual search for faces. *Perception, 23(1),* 95-122. doi: 10.1068/p230095.

Laidlaw, K. E., Risko, E. F., & Kingstonee, A. (2012). A new look at social attention: orienting to the eyes is not (entirely) under volitional control. *Journal of Experimental Psychology: Human Perception and Performance, 38(5),* 1132-1143. doi: 10.1037/a0027075.

Langton, S. R., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition, 107(1),* 330-342. doi: 10.1016/j.cognition.2007.07.012.

Laubrock, J., Engbert, R., & Kliegl, R. (2005). Microsaccade dynamics during covert attention. *Vision Research, 45(6),* 721-730. doi: 10.1016/j.visres.2004.09.029.

Leppänen, J. M. (2016). Using eye tracking to understand infants' attentional bias for faces. *Child Development Perspectives, 10(3),* 161-165. doi: 10.1111/cdep.12180.

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center–periphery organization of human object areas. *Nature Neuroscience, 4(5),* 533-539. doi: 10.1038/87490.

Lewis, M. B., & Edmonds, A. J. (2003). Face detection: Mapping human performance. *Perception, 32(8),* 903-920. doi: 10.1068/p5007.

Lewis, M., & Edmonds, A. (2005). Searching for faces in scrambled scenes. *Visual Cognition, 12(7),* 1309-1336. doi: 10.1080/13506280444000535.

Liu, J., Li, J., Feng, L., Li, L., Tian, J., & Lee, K. (2014). Seeing Jesus in toast: neural and behavioral correlates of face pareidolia. *Cortex, 53,* 60-77. doi: 10.1016/j.cortex.2014.01.013.

Liu-Shuang, J., Ales, J., Rossion, B., & Norcia, A. M. (2015). Separable effects of inversion and contrast-reversal on face detection thresholds and response functions: A sweep VEP study. *Journal of Vision, 15(2),* 11-11. doi: 10.1167/15.2.11.

Looser, C. E., Guntupalli, J. S., & Wheatley, T. (2013). Multivoxel patterns in face-sensitive temporal regions reveal an encoding schema based on detecting life in a face. *Social Cognitive and Affective Neuroscience, 8(7),* 799-805. doi: 10.1093/scan/nss078.

Lyyra, P., Hietanen, J. K., & Astikainen, P. (2014). Anger superiority effect for change detection and change blindness. *Consciousness and Cognition, 30,* 1-12. doi: 10.1016/j.concog.2014.07.013.

Maw, N. N., & Pomplun, M. (2004). Studying human face recognition with the gaze-contingent window technique. *Proceedings of the Annual Meeting of the Cognitive Science Society, 26,* 927-932.

Mayer, K. M., Vuong, Q. C., & Thornton, I. M. (2015). Do people “pop out”? *PLoS One, 10(10),* e0139618. doi: 10.1371/journal.pone.0139618.

Maylott, S. E., Sansone, J. R., Jakobsen, K. V., & Simpson, E. A. (2021). Superior detection of faces in male infants at 2 months. *Child Development, 92(4),* e621-e634. doi: 10.1111/cdev.13543.

McKay, K. T., Grainger, S. A., Coundouris, S. P., Skorich, D. P., Phillips, L. H., & Henry, J. D. (2021). Visual attentional orienting by eye gaze: A meta-analytic review of the gaze-cueing effect. *Psychological Bulletin, 147(12),* 1269-1289. doi: 10.1037/bul0000353.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7(1),* 3. doi: 10.1037/1076-8971.7.1.3.

Meissner, T. W., Prüfer, H., Nordt, M., Semmelmann, K., & Weigelt, S. (2018). Development of face detection in preschool children. *International Journal of Behavioral Development, 42(4),* 439-444. doi: 10.1177/016502541773805.

Melmoth, D. R., Kukkonen, H. T., Mäkelä, P. K., & Rovamo, J. M. (2000). The effect of contrast and size scaling on face perception in foveal and extrafoveal vision. *Investigative Ophthalmology & Visual Science, 41(9)*, 2811-2819.

Menneer, T., Donnelly, N., Godwin, H. J., & Cave, K. R. (2010). High or low target prevalence increases the dual-target cost in visual search. *Journal of Experimental Psychology: Applied, 16(2)*, 133-144. doi: 10.1037/a0019569.

Moscovitch, M., & Moscovitch, D. A. (2000). Super face-inversion effects for isolated internal or external features, and for fractured faces. *Cognitive Neuropsychology, 17(1-3)*, 201-219. doi: 10.1080/026432900380571.

Moulson, M. C., Balas, B., Nelson, C., & Sinha, P. (2011). EEG correlates of categorical and graded face perception. *Neuropsychologia, 49(14)*, 3847-3853. doi: 10.1016/j.neuropsychologia.2011.09.046.

Müller, H. J., & Rabbitt, P. M. (1989). Reflexive and voluntary orienting of visual attention: time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance, 15(2)*, 315-330. doi: 10.1037/0096-1523.15.2.315.

Nachson, I., & Shechory, M. (2002). Effect of inversion on the recognition of external and internal facial features. *Acta Psychologica, 109(3)*, 227-238. doi: 10.1016/S0001-6918(01)00058-0.

Nihei, Y., Minami, T., & Nakauchi, S. (2018). Brain activity related to the judgment of face-likeness: correlation between EEG and face-like evaluation. *Frontiers in Human Neuroscience, 12,* 56. doi: 10.3389/fnhum.2018.00056.

Nuthmann, A., & Canas-Bajo, T. (2022). Visual search in naturalistic scenes from foveal to peripheral vision: A comparison between dynamic and static displays. *Journal of Vision, 22(1),* 10-10. doi: 10.1167/jov.22.1.10.

Oliva, A., & Torralba, A. (2007). The role of context in object recognition. *Trends in Cognitive Sciences, 11(12),* 520-527. doi: 10.1016/j.tics.2007.09.009.

Omer, Y., Sapir, R., Hatuka, Y., & Yovel, G. (2019). What is a face? Critical features for face detection. *Perception, 48(5),* 437-446. doi: 10.1177/0301006619838734.

Or, C. C., & Wilson, H. R. (2010). Face recognition: Are viewpoint and identity processed after face detection? *Vision Research, 50(16),* 1581-1589. doi: 10.1016/j.visres.2010.05.016.

Palanica, A., & Itier, R. J. (2014). Effects of peripheral eccentricity and head orientation on gaze discrimination. *Visual Cognition, 22(9-10),* 1216-1232. doi: 10.1080/13506285.2014.990545.

Palermo, R., & Rhodes, G. (2003). Change detection in the flicker paradigm: Do faces have an advantage? *Visual Cognition, 10(6),* 683-713. doi: 10.1080/13506280344000059.

Peelen, M. V., & Downing, P. E. (2005). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology, 93(1),* 603-608. doi: 10.1152/jn.00513.2004.

Peelen, M. V., & Downing, P. E. (2007). The neural basis of visual body perception. *Nature Reviews Neuroscience, 8(8),* 636-648. doi: 10.1038/nrn2195.

Peelen, M. V., Wiggett, A. J., & Downing, P. E. (2006). Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion. *Neuron, 49(6),* 815-822. doi: 10.1016/j.neuron.2006.02.004.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., & ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195-203. doi: 10.3758/s13428-018-01193-y.

Pereira, E. J., Birmingham, E., & Ristic, J. (2019). Contextually-based social attention diverges across covert and overt measures. *Vision, 3(2),* 29. doi: 10.3390/vision3020029.

Pereira, E. J., Birmingham, E., & Ristic, J. (2020). The eyes do not have it after all? Attention is not automatically biased towards faces and eyes. *Psychological Research, 84,* 1407-1423. doi: 10.1007/s00426-018-1130-4.

Peterson, M. S., Kramer, A. F., & Irwin, D. E. (2004). Covert shifts of attention precede involuntary eye movements. *Perception & Psychophysics, 66(3),* 398-405. doi: 10.3758/BF03194888.

Pongakkasira, K., & Bindemann, M. (2015). The shape of the face template: Geometric distortions of faces and their detection in natural scenes. *Vision Research, 109,* 99-106. doi: 10.1016/j.visres.2015.02.008.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting . *Attention and Performance X: Control of Language Processes, 32*, 531-556.

Prazak, E. R., & Burgund, E. D. (2014). Keeping it real: Recognizing expressions in real compared to schematic faces. *Visual Cognition, 22(5)*, 737-750. doi: 10.1080/13506285.2014.914991.

Prunty, J. E., Jenkins, R., Qarooni, R., & Bindemann, M. (2022). Ingroup and outgroup differences in face detection. *British Journal of Psychology 144(1)*, 94-111. doi: 10.1111/bjop.12588.

Qarooni, R., Prunty, J., Bindemann, M., & Jenkins, R. (2022). Capacity limits in face detection. *Cognition, 228*, 105227. doi: 10.1016/j.cognition.2022.105227.

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading.* Englewood Cliffs, NJ: Prentice Hall.

Rensink, R. A. (2000). Visual search for change: A probe into the nature of attentional processing. *Visual Cognition, 7(1-3)*, 345-376. doi: 10.1080/135062800394847.

Rensink, R. A., O'regan, J. K., & Clark, J. J. (1997). To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science, 8(5)*, 368-373. doi: 10.1111/j.1467-9280.1997.tb00427.

Rieth, C. A., Lee, K., Lui, J., Tian, J., & Huber, D. E. (2011). Faces in the mist: illusory face and letter detection. *i-Perception, 2(5)*, 458-476. doi: 10.1068/i0421.

Ro, T., Friggel, A., & Lavie, N. (2007). Attentional biases for faces and body parts. *Visual Cognition, 15(3),* 322-348. doi: 10.1080/13506280600590434.

Ro, T., Russell, C., & Lavie, N. (2001). Changing faces: A detection advantage in the flicker paradigm. *Psychological Science, 12(1),* 94-99. doi: 10.1111/1467-9280.00317.

Rousselet, G. A., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2005). Spatial scaling factors explain eccentricity effects on face ERPs. *Journal of Vision, 5(10),* 1-1. doi: 10.1167/5.10.1.

Sadr, J., Jarudi, I., & Sinha, P. (2003). The role of eyebrows in face recognition. *Perception, 32(3),* 285-293. doi: 10.1068/p5027.

Sagiv, N., & Bentin, S. (2001). Structural encoding of human and schematic faces: holistic and part-based processes. *Journal of Cognitive Neuroscience, 13(7),* 937-951. doi: 10.1162/089892901753165854.

Santhi, N., & Reeves, A. (2004). The roles of distractor noise and target certainty in search: A signal detection model. *Vision Research, 44(12),* 1235-1256. doi: 10.1016/j.visres.2003.11.011.

Savage, R. A., & Lipp, O. V. (2015). The effect of face inversion on the detection of emotional faces in visual search. *Cognition and Emotion, 29(6),* 972-991. doi: 10.1080/02699931.2014.958981.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience, 25(47),* 11055-11059. doi: 0.1523/JNEUROSCI.2621-05.2005.

Simpson, E. A., Husband, H. L., Yee, K., Fullerton, A., & Jakobsen, K. V. (2014). Visual search efficiency is greater for human faces compared to animal faces. *Experimental Psychology, 61(6)*, 439-456. doi: 10.1027/1618-3169/a000263.

Simpson, E. A., Maylott, S. E., Leonard, K., Lazo, R. J., & Jakobsen, K. V. (2019). Face detection in infants and adults: Effects of orientation and color. *Journal of Experimental Child Psychology, 186*, 17-32. doi: 10.1016/j.jecp.2019.05.001.

Stekelenburg, J. J., & de Gelder, B. (2004). The neural correlates of perceiving human bodies: an ERP study on the body-inversion effect. *Neuroreport, 15(5),* 777-780. doi: 10.1097/01.wnr.0000119730.93564.e8.

Takahashi, K., & Watanabe, K. (2015). Seeing objects as faces enhances object detection. *i-Perception, 6(5)*, doi: 10.1177/2041669515606007.

Tangen, J. M., Murphy, S. C., & Thompson, M. B. (2011). Flashed face distortion effect: Grotesque faces from relative spaces. *Perception, 40(5)*, 628-630. doi: 10.1068/p6968.

Taubert, J., Apthorp, D., Aagten-Murphy, D., & Alais, D. (2011). The role of holistic processing in face perception: Evidence from the face inversion effect. *Vision Research, 51(11),* 1273-1278. doi: 10.1016/j.visres.2011.04.002.

Thierry, G., Pegna, A. J., Dodds, C., Roberts, M., Basan, S., & Downing, P. (2006). An event-related potential component sensitive to images of the human body. *Neuroimage, 32(2),* 871-879. doi: 10.1016/j.neuroimage.2006.03.060.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search. *Psychological Review, 113(4),* 766-786. doi: 10.1037/0033-295X.113.4.766.

Toseeb, U., Keeble, D. R., & Bryant, E. J. (2012). The significance of hair for face recognition. *PloS One, 7(3),* e34144. doi: 10.1371/journal.pone.0034144.

Treisman, A. (1991). Search, similarity, and integration of features between and within dimensions. *Journal of Experimental Psychology: Human Perception and Performance, 17(3),* 652-676. doi: 10.1037/0096-1523.17.3.652.

Tymkiw, M., & Foulsham, T. (2020). Eye tracking, spatial biases and normative spectatorship in museums. *Leonardo, 53(5)*, 542-546. doi: 10.1162/leon_a_01746.

Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79(4),* 471-491. doi: 10.1111/j.2044-8295.1988.tb02747.x.

Van Belle, G., De Graef, P., Verfaillie, K., Busigny, T., & Rossion, B. (2010). Whole not hole: Expert face recognition requires holistic perception. *Neuropsychologia, 48(9),* 2620-2629. doi: 10.1016/j.neuropsychologia.2010.04.034.

Van Belle, G., De Graef, P., Verfaillie, K., Rossion, B., & Lefevre, P. (2010). Face inversion impairs holistic perception: Evidence from gaze-contingent stimulation. *Journal of Vision, 10(5),* 10-10. doi: 10.1167/10.5.10.

Van Belle, G., Lefèvre, P., & Rossion, B. (2015). Face inversion and acquired prosopagnosia reduce the size of the perceptual field of view. *Cognition, 136,* 403-408. doi: 10.1016/j.cognition.2014.11.037.

Van Diepen, P., & d'Ydewalle, G. (2003). Early peripheral and foveal processing in fixations during scene perception. *Visual Cognition, 10(1),* 79-100. doi: 10.1080/713756668.

Viola Macchi, C., Turati, C., & Simion, F. (2004). Can a nonspecific bias toward top-heavy patterns explain newborns' face preference? *Psychological Science, 15(6),* 379-383. doi: 10.1111/j.0956-7976.2004.00688.x.

Wang, W., Miao, D., & Zhao, L. (2014). Automatic detection of orientation changes of faces versus non-face objects: a visual MMN study. *Biological Psychology, 100,* 71-78. doi: 10.1016/j.biopsycho.2014.05.004.

Wardle, S. G., Taubert, J., Teichmann, L., & Baker, C. I. (2020). Rapid and dynamic processing of face pareidolia in the human brain. *Nature Communications, 11(1),* 4518. doi: 10.1038/s41467-020-18325-8.

Wardle, S., Seymour, K., & Taubert, J. (2017). Decoding face pareidolia in the human brain with fMRI. *Journal of Vision, 17(10),* 294-294. doi: 10.1167/17.10.294.

Wegrzyn, M., Vogt, M., Kireclioglu, B., Schneider, J., & Kissler, J. (2017). Mapping the emotional face. How individual face parts contribute to successful emotion recognition. *PloS One, 12(5),* e0177239. doi: 10.1371/journal.pone.0177239.

Westheimer, G. (1965). Visual acuity. *Annual Review of Psychology, 16(1),* 359-380.

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided search: an alternative to the feature integration model for visual search. *Journal of Experimental Psychology: Human Perception and Performance, 15(3),* 419-433. doi: 10.1037/0096-1523.15.3.419.

Wolfe, J. M., Horowitz, T. S., Van Wert, M. J., Kenner, N. M., Place, S. S., & Kibbi, N. (2007). Low target prevalence is a stubborn source of errors in visual search tasks. *Journal of Experimental Psychology: General, 136(4),* 623-638. doi: 10.1037/0096-3445.136.4.623.

Yang, C. T., Shih, C. H., Cheng, M., & Yeh, Y. Y. (2009). Similarity modulates the face-capturing effect in change detection. *Visual Cognition, 17(4),* 484-499. doi: 10.1080/13506280701822991.

Yin, R. K. (1969). Looking at upside-down faces. *Journal of Experimental Psychology, 81(1),* 141-145. doi: 10.1037/h0027474.

Yip, A. W., & Sinha, P. (2002). Contribution of color to face recognition. *Perception, 31(8),* 995-1003. doi: 10.1068/p3376.

Young, A. W., Hay, D. C., McWeeny, K. H., Flude, B. M., & Ellis, A. W. (1985). Matching familiar and unfamiliar faces on internal and external features. *Perception, 14(6),* 737-746. doi: 10.1068/p140737.

Zhou, L. F., & Meng, M. (2020). Do you see the “face”? Individual differences in face pareidolia. *Journal of Pacific Rim Psychology, 14,* e2. doi: 10.1017/prp.2019.27.