Capacity Limits in Face Detection

Rana Qarooni

Doctor of Philosophy in Psychology

University of York
Psychology

September 2022

Abstract

Our ability to socialise and interact with others is underpinned by our ability to extract social information from faces. However, these faces must first be detected from the complex visual environment before any social information is extracted. Prior face perception work has focused mainly on later stages of face processing. The early stage of face detection has been relatively neglected. As such, some basic questions in face detection are outstanding. Can we know that more than one face is present? How many faces can be detected at once? How do viewing conditions affect performance? This thesis adapted paradigms from face perception and numerical cognition research to address these questions. First, Chapter 2 used a 'subitizing of faces' approach to compare multiple target detection for faces and other types of stimuli. A detection advantage was found for faces over non-face stimuli, irrespective of face inversion. Chapter 3 used a novel 'fixed/mixed' judgment task to test for capacity limits in face detection. The findings supported an efficient parallel detection mechanism for multiple faces. Chapter 4 adapted search tasks to assess how multiple face detection is affected by different viewing conditions, including aspects of the task and presentations. The findings in this chapter indicate a typical face detection span of four faces, plus or minus one. They also show how the visual complexity and meaningfulness of the surrounding scene affect performance. This thesis establishes that detecting multiple faces in complex visual scenes is an efficient parallel process for up to four faces. It also contributes several methodological innovations that can be adapted to address related research questions.

Table of Contents

List of Figures

Acknowledgements

On one of the first few days of my PhD, I was having a conversation with another member of the department about my project and who I am working with. When I mentioned that I am working with Rob Jenkins, I was told that I am lucky. I am lucky because some supervisors are intelligent and ‘know-what-they’re-doing’, and others are friendly. But Rob is that rare kind of supervisor who far exceeds that combination. This has been proven to me every day since, with every question and idea I presented to Rob, he has offered invaluable guidance, support, and mentorship, allowing me to explore while also steering me with consideration and thoughtful insight. So to Rob, a simple thank you is not enough to express how grateful I am to you for helping me see the ‘frame’ when all I could see were the separate pieces (and for patiently listening to me while I slowly try to put words together into coherent sentences throughout our meetings).

I would also like to thank my Thesis Advisory Panel, Dr Silke Göbel and Dr Cade McCall for their comments and insight throughout my research. Silke’s help in navigating the numerical cognition literature at the beginning of my first year was foundational for the rest of my research. Also my sincere thanks to the Face Lab Group for listening to my presentations and sharing their significant expertise and advice. I would like to specifically thank Dr Mila Mileva whose help, advice, and guidance while I tried to learn MATLAB and coding from scratch is something I will always be grateful for.

Also thank you to my parents for their unfailing love and support. I am grateful for everything you have done for me. Without your encouragement and belief in me, I would not be where I am today. Even though the pandemic has grounded me in Bahrain for a year and a half, I was glad to have that time to be back home again with you, Ahmed, and Ala’a (+ cats). I would also like to acknowledge the impact my cousins and my extended family have had on my PhD journey. These past few years may have been horrible, but I am so thankful that we were all together to deal with it in our usual hilarious ways.

This short acknowledgement section does not do justice in recognising the many wonderful and amazing people who have impacted my PhD journey. I have been privileged enough to have made many amazing, thoughtful, considerate, intelligent, and genuinely funny friends throughout my PhD. So to Lamisse, I am so happy to have meet you and to be able to call you a lifelong friend – finally someone who shares my opinions about food. To Fed, Lydia, Kirri, Bee, and Jo thank you for being the most fun, caring and supportive group of friends I made in the UK. My third year would have been a lot less entertaining without you all.

Finally, I would like to acknowledge my friends back home. It may be a cliché, but you all have been with me since the beginning. Thank you for always piloting my experiments and for listening to me go on and on about my research. Thank you to Maryam for the tough love and for always being there for me when I need you – and also for *always* knowing when to call. Thank you as well to Reem, Maha, Naziha, Hala, and Sara for always finding the time to catch up no matter where we are in the world. I would also like to give a special thank you to Feras who has spent countless hours – despite the massive time difference between here and the US – helping me code and debug my experiments, and for patiently explaining to me every single line of a code or function when I insist on. You gave me the tools to understand the logic behind the coding errors rather than simply fixing them for me, and for that I will always be grateful.

Author’s Declaration

I, Rana Qarooni, declare that this thesis is a presentation of original work and that I am the sole author under normal terms of supervision by Professor Rob Jenkins. This work has not been previously presented at this or any other university. All sources are acknowledged as References. This research was funded by a research grant from the Leverhulme Trust (RPG-2019-085) to Markus Bindemann and Rob Jenkins.

Chapter 3 of this thesis has been published in the following peer-reviewed journal. Its formatting has been adapted to fit the formatting requirements of this thesis.

> Qarooni, R., Prunty, J., Bindemann, M., & Jenkins, R. (2022). Capacity limits in face detection. *Cognition*, *228*, 105227.
>
> Rana Qarooni: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft. Jonathan Prunty: Conceptualization, Writing – review & editing. Markus Bindemann: Conceptualization, Writing – review & editing. Rob Jenkins: Conceptualization, Methodology, Supervision, Writing – review & editing.

Chapter 4 of this thesis has been submitted for publication. Its formatting has been adapted to fit the formatting requirements of this thesis.

> Qarooni, R., Prunty, J., Bindemann, M., & Jenkins, R. (2022). Multiple Face Detection. *(Submitted)*.
>
> Rana Qarooni: Conceptualization, Methodology, Formal analysis, Data curation, Writing – original draft. Jonathan Prunty: Conceptualization, Writing – review & editing. Markus Bindemann: Conceptualization, Writing – review & editing. Rob Jenkins: Conceptualization, Methodology, Supervision, Writing – review & editing.

Supervisor’s Declaration

I, Rob Jenkins, am listed as a co-author on two empirical papers which make up two chapters of this thesis. In each of the reported studies, the work is primarily that of Rana Qarooni. For each paper, Rana compiled the relevant stimuli, created and coded the experimental design, and completed all of the data collection and data analysis. Rana wrote the first draft of each paper.

Prof Rob Jenkins

Chapter 1 – General Introduction

Face perception can be understood as a two-stage process. An initial detection stage registers the presence of a face in the environment, and a subsequent analysis stage extracts meaning from the face. The sheer number of faces we encounter in our social lives raises the question of bandwidth along this processing pipeline. Can we process multiple faces in parallel, or are we constrained to processing single faces serially? The current thesis addresses this question for the initial stage of face detection.

To date, most psychological research on face perception has concentrated on later processes, such as analysis of emotional expression or personal identity (for reviews, see Bruce & Young, 2013; Calder, 2011; Freiwald et al., 2016; O'Toole, 2005; O'Toole & Castillo, 2021; Valentine, 1991). As such, the initial detection stage remains comparatively underexplored as a topic in cognition. Most face detection literature stems from computer vision research concerned with developing automatic face detection algorithms (Sinha et al., 2006). These algorithms identify candidate face regions in an image so that processing resources can be allocated appropriately (Hjelmås & Low, 2001; Sinha et al., 2006; Wang & He, 2019). This task decomposition echoes the distinction between face detection and face processing in humans (Lewis & Ellis, 2003; Tsao & Livingstone, 2008).

Interestingly, many face detection algorithms claim to be inspired by psychological and physiological findings (Sinha, 2002; Sinha et al., 2006). Yet the volume of automatic face detection research far exceeds the volume of underpinning psychological research. This dynamic suggests a potential for impact beyond experimental psychology. Further research into face detection by humans will not only deepen our understanding of visual cognition but also broaden the evidence base for bio-inspired computer vision allowing for meaningful comparisons of human and machine systems.

This General Introduction is divided into three sections. The first section reviews the psychological literature on face detection and argues that a quantitative understanding of face detection is missing. The second section draws on concepts from numerical cognition research and argues that these concepts could be usefully adapted for face detection research. The final section provides an overview of the experimental chapters in this thesis, setting out the central research questions and the methods used to address them.

1.1 Face Detection: A Review of the Current Evidence

Face detection involves registering the presence of faces by comparing regions of the visual environment to a stored face template. In this view, detection is a separate prerequisite step that selects appropriate stimuli for subsequent analysis (Lewis & Ellis, 2003; Robertson et al., 2017; Tsao & Livingstone, 2008). Although a few pioneering researchers have examined face detection directly (Bindemann & Burton, 2009; Burton & Bindemann, 2009; Lewis & Edmonds, 2003, 2005; Nothdurft, 1993; Purcell & Stewart, 1986, 1988), much of our current understanding is gleaned from studies of related processes, including face identification and visual search. The following section provides an overview of the current face detection evidence, discussing both the Face Detection Effect (FDE) for upright faces over other inverted faces as observed by (Purcell & Stewart, 1986, 1988) and the general advantage over non-face stimuli which is subsequently termed as the 'Face Detection Advantage'. An overview of the role of face detection within face processing will also be presented. This section shall then discuss *qualitative* aspects of the face detection template and its properties before turning to the limited research on *quantitative* aspects of face detection.

1.1.1 Detection as a Prerequisite for Face Recognition

Several strands of research have been used to make the case that faces, specifically in their upright orientation, are 'special' stimuli for the perception system (Farah et al., 1998; Kanwisher et al., 1997; Tanaka & Farah, 2007; Yin, 1969a). In a similar vein, early work on face detection suggested that detecting a single upright face differs from detecting other types of face stimuli. Purcell & Stewart (1986) reported a Face Detection Effect (FDE) whereby intact photographs of faces needed a shorter presentation time to be detected than when the internal configuration of eyes, nose, and mouth were arbitrarily jumbled. In a follow-up study, upright schematic faces were also detected at shorter presentation times than inverted faces (Figure 1.1; Purcell & Stewart, 1988). Manipulating exposure duration times is seemingly an effective way to limit face processing to detection alone, as the task demands no further action from the participant once a face is detected. However, the forced-choice task used by Purcell & Stewart (1986, 1988) was more akin to a visual search task than a detection task, as the participants had to indicate the presence or absence of a single target stimulus (Lewis & Ellis, 2003). Nonetheless, the FDE shows an advantage for detecting faces with the correct internal configuration over rearranged stimuli that are otherwise visually similar.

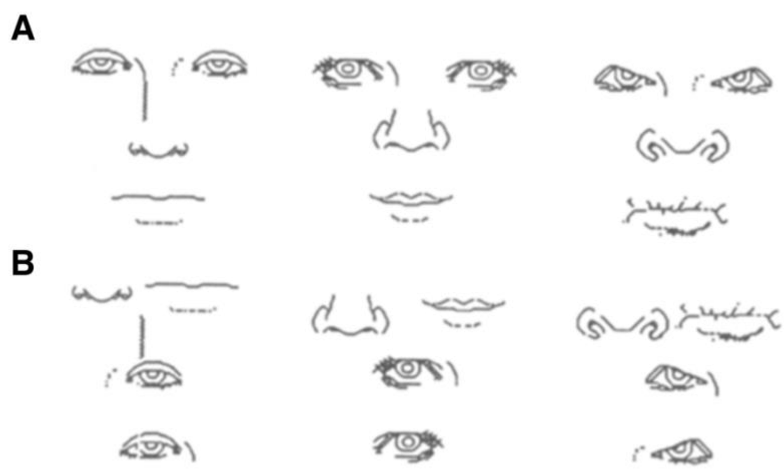

Figure 1.1 Intact (A) and jumbled (B) faces used in experiments 6 in Purcell & Stewart (1988).

Face detection studies also point to a face advantage and a hierarchy in detection efficiency for different visual objects. Converging evidence from eye-tracking and visual search studies show that a face can be detected more efficiently than a face-like object (such as a pareidolic or illusory face seen in Figure 1.2), which can, in turn, be detected more efficiently than a non-face object (Crouzet et al., 2010; Keys et al., 2021; Purcell & Stewart, 1986, 1988; Wardle et al., 2020). This enhanced detection performance also appears to emerge in infancy and be specific to human faces rather than generalising to other mammal or animal faces (Simpson, Buchin, et al., 2014; Simpson, Husband, et al., 2014; Simpson, Maylott, Leonard, et al., 2019; Simpson, Maylott, Mitsven, et al., 2019). The preferential detection of faces over other stimuli echoes findings from face perception literature on infant looking times (Di Giorgio et al., 2012; Farroni et al., 2002, 2005, 2007; Simion et al., 2005). This finding supports a detection mechanism specialised for detecting a face over another item and points to some level of specificity in the face detection template.

Figure 1.2 Pareidolic faces (top row), matched non-face objects (middle row), and real faces (bottom row) used by Keys et al. (2021).

A similar face detection advantage is also seen in a series of eye-tracking studies by Crouzet et al. (2010). When two images (a face and a vehicle) were simultaneously presented in the left and right visual fields, automatic eye saccades towards faces occurred as early as 100 – 110 ms. These automatic eye saccades seem outside of conscious control as they were biased towards faces even when participants were instructed to look at the vehicle. Crouzet et al. (2010) and later Crouzet & Thorpe (2011) argue that the presence of a face can be determined within 80 ms, given that around 20 ms is needed to initiate eye movements. These ultra-rapid eye saccade findings show that the status of faces as 'special stimuli' is possibly present even at the detection stage — not just the later stage of extracting meaning from the face. When these findings are considered alongside the FDE, they suggest the presence of a specialised detection mechanism that is tuned to human faces.

While faces may be preferentially detected over non-face stimuli, the precise tuning of the face detection template — what counts as a face to the face detection system — remains an open question. Sensitivity to orientation is interesting in this regard. It is well established that familiar faces are easier to identify when upright than when inverted (Farah et al., 1995, 1998; Tanaka & Farah, 2007). That is, face identification is orientation sensitive. However, upright and inverted faces appear equally potent in biasing spatial attention in a cueing task (Bindemann & Burton, 2008). Consequently, face detection appears to be less sensitive than identification to stimulus orientation. The authors attribute this observation to shared featural qualities between upright and inverted faces. Similar findings emerged in saccadic eye movement studies. While the aforementioned Crouzet et al. (2010) study found automatic eye saccades for faces over vehicles, it did not incorporate faces in an inverted orientation. However, in other saccadic eye studies by Devue et al. (2012) and Laidlaw et al. (2015), upright and inverted faces were compared directly and found to bias saccadic eye movements and trajectories over a non-face target stimulus. This bias was present for upright and inverted faces but not scrambled faces or animate objects, suggesting that the detection template may be tuned to properties that are common to upright and inverted faces. This finding further supports specificity in the face detection template, but not sensitivity.

Whatever the precise tuning of the face detection mechanism, the findings reviewed above suggest that faces may be in some sense ‘special’ even at the detection stage. However, evidence of ‘specialness’ in face detection, as well as later face processes, does not imply unity of face perception. Several findings suggest that face detection is dissociable from identification. Robertson, Jenkins, & Burton, (2017) operationalised face detection as participants’ tendency to see faces in pareidolia images and photos of clouds. They also measured detection more directly through participants’ ability to locate a face image in a cluttered scene. Detection performance in these tasks showed no relationship to identification performance, as measured using a standard face matching task. Fysh (2018) replicated and extended these findings. In addition to different detection tasks and more challenging matching tasks, face memory tasks were included to measure face recognition. Face detection did not correlate with either face recognition (memory) or identification (perception), but the latter two processes were strongly correlated with each other. These observations all support the notion that face detection is separable from recognition and identification.

Converging evidence from cognitive neuroscience also points to face detection as a separable process. Using magnetoencephalography (MEG), Liu, Harris, & Kanwisher (2002) found that face-selective M100 responses at 100 ms after face presentation correlated with behavioural performance in a face categorisation task but not face recognition. A later M170 response 170 ms post-stimulus correlated with both categorisation and recognition performance. A face categorisation task in which stimuli are presented at fixation does not require localising the face in a visual environment (Bindemann & Lewis, 2013a). Nonetheless, when these MEG results are considered alongside the aforementioned behavioural findings, they indicate an ordered feed-forward progression in face perception whereby faces are first detected and then subject to deeper processing.

However, evidence from neuropsychological studies points to a more nuanced division between detection and later face processes. Le Grand et al., (2006) and Garrido, Duchaine, & Nakayama (2008) report prosopagnosia patients with the expected deficits in recognition and identification, but some showed intact detection while others showed impaired detection. This calls into question whether impairments in recognition and identification sometimes stem from detection. In a feed-forward model of face perception, deficits in the early stages (i.e. detection) should impair later stages. However, if face detection was completely separable from later stages, then impairments in recognition and identification would be independent of impairments in detection, and vice versa. Xu & Biederman (2014) examined this issue directly with patient MJH who had bilateral lesions in the temporal-occipital cortices, including the fusiform and occipital face areas. Previously MJH was reported to have normal, or near-normal face detection but impaired face individuation. Xu & Biederman (2014) used a different detection paradigm from the experiments described above to further investigate detection abilities in MJH. Images of faces and cars were Fourier-transformed and subjected to visual noise—a means of manipulating detectability without affecting luminance or contrast. Participants were then presented with a face and a car on either side of fixation and asked to indicate which image was a face. Compared to controls, MJH was impaired on this detection task, as well as face individuation tasks. The researchers argue that, at least on a neurological level, there is some overlap between the detection and identification of faces. This finding further supports the proposed feed-forward model of face perception whereby detection leads into later processes, as well as indicates that these different stages are not completely independent of each other.

Investigating capacity limits could shed further light on the dissociation between face detection and later stages of face perception. Bindemann, Jenkins, &

Burton (2007) propose that identification is capacity limited such that only one face can be identified at a time. In instances where a non-target face flanked a target face, only the target was identified; but the second flanker face was also attended to and detected. Bindemann et al. (2007) refer to this as a bottleneck in identification, arguing that when a face is being identified, it occupies all identification processing resources. A second face can still be detected; however, it cannot move to the identification stage when that is already occupied by the preceding face. This interpretation supports the notion of face detection as a separate prerequisite step to identification. However, it also raises the question of whether the bottleneck applies at the detection stage itself, or after it. One interpretation could be that detection is also capacity limited to one face. This single face then moves on to identification, which is also capacity-limited, freeing up its space in the detection stage and allowing for a single new face to occupy it. An equally plausible alternative interpretation could be that the detection stage is not strictly capacity limited to one face at a time. Instead, more than one face could be detected, but the bottleneck at the later face processes, as suggested by Bindemann et al. (2007), constraints face perception. These alternative hypotheses are presented visually in Figure 1.3. The current results and existing literature on face detection do not uphold one interpretation over the other. Directly testing for capacity limits at the detection stage should resolve this issue. However, doing so requires appropriate experimental methods to be developed.

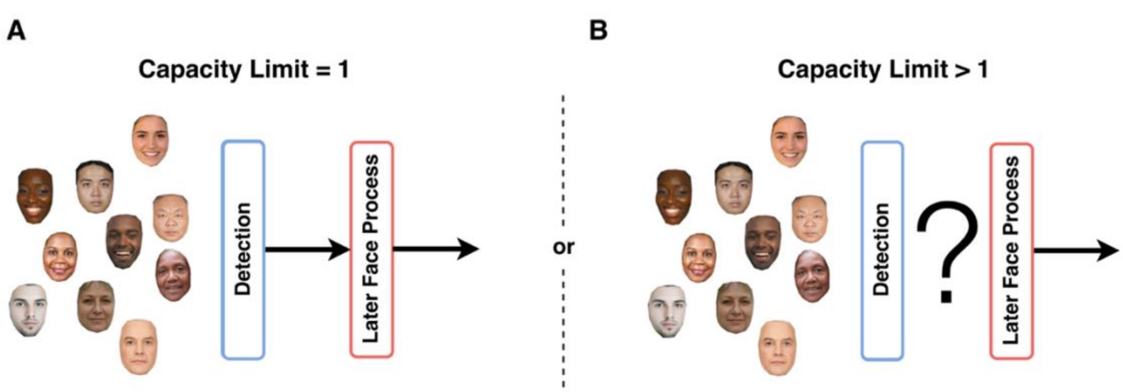

Figure 1.3 Alternative hypotheses accounting for the bottleneck of face processing if (A) detection was limited to one face at a time, or (B) more than one face at a time.

Most of the current face detection research uses variations of visual search tasks or face categorisation tasks to assess detection. Face categorisation tasks present participants with a stimulus near fixation on a blank background and instruct them to report if the stimulus is a face or not (Lewis & Ellis, 2003). These tasks can provide insight into how efficiently category membership is established for faces (that is, knowing that the stimulus is a face). On the other hand, visual search tasks have often been helpful in the localisation of faces (that is, finding the specified target in a visual scene). Visual search tasks were proposed by Treisman & Gelade (1980) to explore serial-*vs*-parallel processing. In these tasks, participants are required to search for a single target face among a varying number of distractors and then report whether the face is present or absent. Target detection time is taken as a proxy for serial or parallel processing. If detection time increases as the number of display items (set size) increases, then the search process is deemed serial as it requires each item to be checked in turn. However, if detection time remains relatively constant over set size, the search is considered parallel, as the target 'pops out' from an arbitrary number of distractors. Wolfe (1994) built upon this theoretical framework, suggesting that when targets pop out under a parallel search, it may be due to shared underlying properties being detected preattentively. On this view, visual search tasks can be very useful for understanding face detection. By distinguishing serial and parallel processes for different target–distractor combinations, we can elucidate properties of the cognitive face template.

However, in the context of *multiple* face detection, categorisation and visual search tasks are limited in how informative they can be. Face categorisation tasks, which usually present a single face at fixation, eliminate the need to assess the visual environment for candidate face regions. While the distinction made between targets and distractors in visual search tasks could potentially prime and bias detection towards the target. Furthermore, visual search tasks necessitate a self-terminating scanning procedure where the participant must search all of the visual environment for a target. This active search seems at odds with the passive nature of face detection outside of the laboratory. In everyday social life, a viewer may readily detect a face in the environment, even while daydreaming about unrelated matters. Moreover, the self-terminating nature of standard visual search tasks requires the participant to respond as soon as one face is found, whether or not additional faces are also present. This task design makes standard visual search unsuitable for assessing multiple face detection. The final section of the General Introduction expands on task design considerations.

1.1.2 Serial-vs-Parallel Processing in Face Detection

Nothdurft (1993) was among the first to explore face detection using visual search tasks. In a series of experiments, target faces were embedded among distractor faces that were either inverted, jumbled or varied in emotional expression (see Figure 1.4). In each presentation, the target was equally likely to be present or absent, and set size varied up to 48 items. Across all experiments, increases in distractor and size led to increases in target detection times, suggesting a serial search process. Kuhn & Jolicoeur (1994) extended these findings, adding skin tone and hair colour and finding further evidence for serial processing in face detection. However, both of these experiments used schematic drawn faces that were the basic properties of a face but were not real faces.

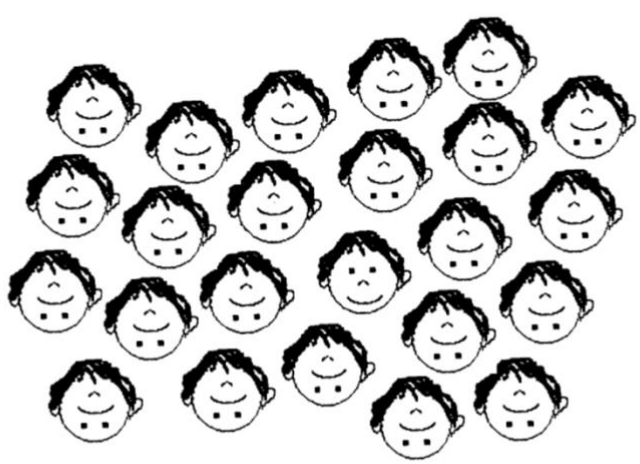

Figure 1.4 Upright and inverted schematic face displays used in experiment 7 of Nothdurft (1993). Visual search times for the target upright face increased as inverted face distractor set size increased.

When naturalistic visual environments with real faces were used to explore face detection, the process was found to be parallel rather than serial. Lewis & Edmonds (2005) ran a series of experiments where stills from a TV soap opera were divided into different-sized arrays and presented in colour (Figure 1.5). A single target (an upright face) could be either present or absent. Distractors varied across experiments and were either intact or scrambled natural backgrounds taken from the same scene, or inverted faces. Face detection was similarly efficient when distractors were intact or scrambled scenes, and each additional distractor added less than 10 ms to response times. This very shallow search slope was taken as evidence of parallel processing in which the target face pops out no matter the size of the visual array. However, the search slope indicated serial search process when upright faces were embedded among inverted faces (~20 ms per item). This pattern accords with the serial search findings of Northduft (1993) but seems at odds with Purcell and Stewart (1986; 1988) and Crouzet et al. (2010) on the upright face advantage in detection. Lewis and Edmonds (2005) argue that this pattern of results is due to the high similarity between upright and inverted faces. Target–distractor similarity is known to influence visual search task results (Duncan & Humphreys, 1989). The authors, therefore, propose that face detection proceeds in stages. A first stage involves extracting all common visual properties of the face, while a second stage deals with the rotation and matching of faces in the upright configuration. This account would explain how a single upright face embedded in a naturalistic environment is detected in a rapid manner. However, when a single face is embedded among inverted faces, the high level of visual similarity means that all the items are initially detected at once, but then a cost is incurred for extracting the upright face. This staged interpretation of face detection could resolve the discrepancy between the upright face advantage found by previous studies and the exception to the inversion effect in detection found by Bindemann & Burton (2008).

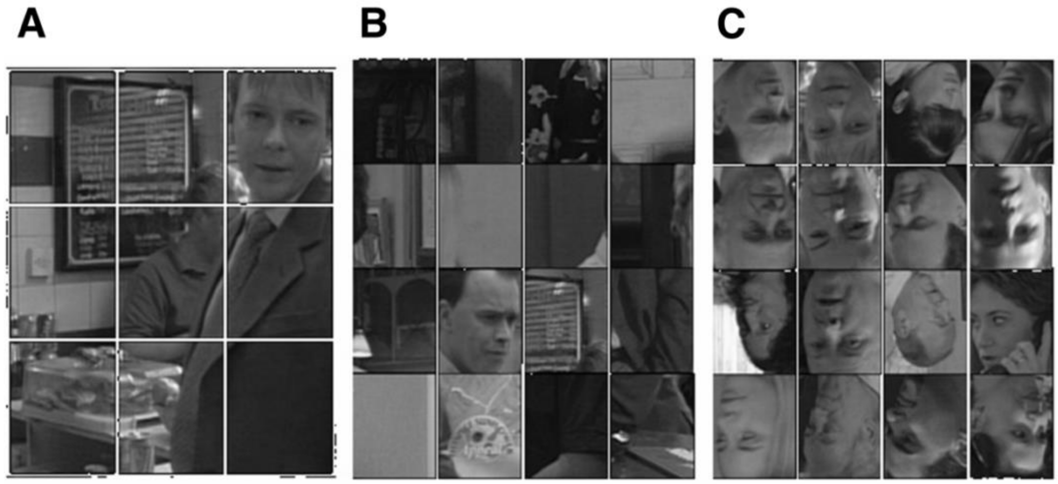

Figure 1.5 Example stimuli from Lewis and Edmonds (2005) including the unscrambled source image (A), the 4 × 4 scrambled scenes with an upright face target used in experiment 1 (B), and the 4 x 4 displays containing an upright face target among inverted face distractors used in experiment 2. Actual stimulus displays were in colour and of a higher quality then the current reproductions.

Lewis and Edmonds (2005) conducted further studies to explore the visual properties of faces implicated in the initial parallel detection stage. The same paradigm as their previous experiments was used, but images were manipulated so that colour was removed or reversed, the image was blurred, or luminance was reversed. In the first three cases, the results were the same as non-manipulated images. However, when luminance was reversed, parallel detection was compromised. Luminance reversal disrupts the veridical lighting pattern, producing images that resemble photographic negatives. Lewis and Edmonds (2005) suggest that detection initially picks up on common luminance patterns that are shared by all faces, no matter the orientation. Orientation then becomes important at the second stage of detection, depending on the task. However, the researchers do not entirely discount a role for colour in face detection.

The role of colour was directly investigated by Bindemann & Burton (2009). They found that face detection performance was impaired when faces were presented in greyscale, whether or not the background visual environment was coloured.

Detection was also impaired when faces were hue reversed and rendered in unnatural colours (e.g. blue skin). This finding suggests that it is the loss of skin colour specifically that affects detection, rather than the loss of colour in general. Bindemann & Burton (2009) investigated this possibility further by retaining skin colour in only half of the face, with the other half in greyscale. In this condition, detection was impaired, suggesting that colour information and aspect ratio information both contribute. Pongakkasira & Bindemann (2015) further explored the role of aspect ratio by manipulating face height-to-width ratio. Unlike familiar face identification (Hole et al., 2002) face detection was impaired when shape was distorted. This finding underscores the importance of aspect ratio information in face detection. It also underscores the dissociation between detection and identification processes.

Each of the studies mentioned above that used visual search tasks using naturalistic visual environments investigated a different aspect of face detection. However, they all seem to support the involvement of a template-matching procedure that lies on lower-level visual properties shared by upright and inverted faces, such as luminance, as well as general shape and colour cues.

Taken together, the available evidence usefully constrains how face detection in humans could work. For example, it points to an important role for the local structure of luminance in the stimulus and appears to rule out a key role for global orientation. Even so, given the overall popularity of face perception as a topic in psychology, it is surprising how few studies have explored face detection directly. Fewer yet have examined face detection in naturalistic viewing conditions. The upshot is that basic questions in face detection remain unanswered. For example, it is not known whether humans can detect multiple faces concurrently. Given that humans evolved in social groups, and often encounter people in groups (Dunbar, 1998, 2012; Zhou et al., 2005), it seems plausible that multiple face detection would be an asset. On the other hand, later processes in face perception (e.g. extraction of personal identity, semantic information, eye direction) appear to be strictly capacity limited to one face at a time (Bindemann et al., 2007; Bindemann, Burton, et al., 2005). Behavioural experiments could resolve this issue by presenting multiple faces in the visual environment in the context of a face detection task.

1.2 Visual Item Enumeration: Evidence from Numerical Cognition

Visual understanding of numerosity has been a topic in numerical cognition research for several decades. Numerical cognition methods offer a way to explore capacity limits in face detection and to investigate the maximum number of faces that can be detected. These paradigms were originally designed to assess how items in the visual environment are enumerated and represented non-symbolically. However, they can be adapted to evaluate multiple face detection in the visual environment. Doing so could shed light on whether face detection is capacity-free (and operates in parallel), or capacity-limited (and operates serially); and if there is a capacity limit, to quantify that limit: how many faces can we detect at once? If only one face can be detected at a time, we can infer a serial process. If all the faces in a visual scene can be detected simultaneously, we can infer a parallel process. The following section outlines theories and paradigms from numerical cognition literature that assess the visual item enumeration of small and large quantities.

1.2.1 Subitizing: Small Exact Sense

Subitizing refers to the rapid and accurate enumeration of multiple items in the visual environment (Kaufman & Lord, 1949). Typical subitizing paradigms involve 1-8 dots rapidly presented on the screen for a very brief duration. Participants are tasked with responding as quickly and accurately as possible, and both accuracy and reaction times are measured. Typically, up to 3 or 4 items can be rapidly and accurately enumerated but with 5 or more items, accuracy decreases, and reaction times increase. Importantly, the reaction time difference between the subitizing of 1-3 or 4 items is typically within a 100 ms range, and accuracy is usually at ceiling (Piazza et al., 2011). This observation is taken as support for rapid and parallel processing for the visual enumeration of small quantities. Subitizing has been replicated using bars instead of dots (Egeth et al., 1988), in the tactile domain as well as the visual domain (Riggs et al., 2006), in children as well as adults (albeit with slower overall reaction times; Chi & Klahr, 1975), and even in chimpanzees (Murofushi, 1997).

Subitizing differs from counting or highly accurate estimation not only behaviourally, but also in terms of distinct brain regions involved (Choo & Franconeri, 2014; Demeyere et al., 2012). Unlike counting, subitizing is considered a capacity-limited preattentive process of visual item individuation. This theory stems from vision research and, in its most popular form, is termed 'FINgers on INSTanitation' or FINSTs (Trick & Pylyshyn, 1994). Trick & Pylyshyn's (1994) FINSTs theory suggests the presence of a domain-general capacity-limited process of item individuation, in which up to four mental indices can be occupied parallelly to track objects in the visual environment. Up to 3 or 4 FINSTs can be assigned at once, accounting for the high accuracy and short reaction time for up to 3 or 4 items in the visual environment. In some respects, subitizing is similar to detection. In both cases, particular visual objects must be segmented from their surroundings. In subitizing, these objects are merely enumerated, but in face detection, they may progress to further face processing.

Subitizing has often been considered a preattentive process (for a review, see (Gilmore et al., 2018b; Katzin et al., 2019), but later studies have found that subitizing can be affected by attention and visual load and visual working memory (WM) (Alvarez & Cavanagh, 2004; Cavanagh & Alvarez, 2005; Cowan, 2001; Eayrs & Lavie, 2018; Eayrs & Lavie, 2021; Piazza et al., 2011; Railo et al., 2008). Eayrs & Lavie's (2021) recent work applying load theory (the influence of the perceptual load on attention; Lavie, 1995) to subitizing lends support to this claim. The authors found that when visual memory was overloaded by distractor quantities within the subitizing range, visual item enumeration was distributed, but the effect disappeared beyond the subitizing range. Moreover, individual differences in WM were correlated with subitizing performance, but there was evidence to support a general capacity in visual perception (Eayrs & Lavie, 2018; Eayrs & Lavie, 2021). Piazza et al. (2011) agree that subitizing reflects a parallel process of visual item individuation, making it a viable method and paradigm to investigate capacity limits in face detection.

1.2.2 Approximate Number Sense (ANS): Large Approximate Sense

From a subitizing perspective, it is plausible that face detection is capacity limited to a small number of faces. However, it may be likely that larger quantities of faces can be detected. The approximate number sense is thought to be involved in instances where the number of items in the visual environment is too large for subitizing, and counting is too effortful. The ANS is a separate process from subitizing and is responsible for the imprecise estimation of large quantities of items in the visual environment (Feigenson et al., 2004). It gives an indication of the intuitive non-symbolic mental representation of these large quantities. Consequently, from an ANS perspective, it is plausible that we have the capacity to detect and extract the presence of many faces, but we do so with less precision.

Non-symbolic comparison tasks are used to measure the ANS. Arrays of dots are presented on the left and right sides of the display, and the ratio of the number of dots is manipulated. Participants are tasked with selecting the array they perceive as being the largest, and accuracy is measured. The ratio difference is thought to drive performance rather than absolute differences between arrays such that the closer the ratios are to 1:1, the less accurate the ANS and performance are. However, when participants are additionally tasked with estimating the quantities presented, they show large variations in estimates as array size increases but tend to usually underestimate the actual quantities (Gilmore et al., 2011; M. Guillaume et al., 2013; Izard & Dehaene, 2008; Lyons & Beilock, 2011).

Accuracy and performance in these non-symbolic tasks are used to create a mental number line (MNL). The MNL represents the internal mental activation and representation of various large quantities. In the MNL, activation for large quantities is imprecise and overlaps with other nearby quantities making them harder to separate. This is akin to the large variations in estimation as array size increases. For instance, when participants are tasked with estimating the quantity of 200 dots, they may provide estimates of between 150-300 due to overlap in the activation of these quantities on the MNL. However, if participants are tasked with estimating an array of 20 dots, the estimates, they may provide might range between 15-25 dots (Gilmore et al., 2018a; Izard & Dehaene, 2008).

One of the most important aspects of the ANS is its abstractness, it can be used for comparison, addition, and subtraction within and across format and modality (Barth et al., 2003). The ANS is thought to tap into this rapid intuitive mental representation of quantities. But it also demonstrates sensitivity to lower-level visual factors such as luminance or colour as well as other factors such as surface area and dot size. This may be due to the high volume of ANS studies in the visual domain. However, Gebuis, Cohen Kadosh, & Gevers (2016) propose an alternative explanation. While this explanation has not been extensively tested, it suggests that the ability to process large quantities of items is actually underlined by a sensory integration system that manifests as an ANS. Essentially the way items are enumerated in the visual environment is dependent on the common sensory properties of stimuli. In the case of face detection, this could be the luminance, colour, and aspect ratio information thought to be involved in template matching.

The ability of the ANS to generalise across common visual properties shared by stimuli in the visual environment implicates it as a useful mechanism for detecting large quantities of faces (e.g. crowds, demonstrations, large gatherings etc). If we have the capacity to register the presence of numerous faces, even if the quantity is approximate, it may point to a further distinction from later face processes that are capacity limited to one. However, by virtue of being approximate, the ANS does not easily lend itself to assessing the detection cost per additional face and, in turn, the capacity limits of face detection. Moreover, as later face processes are capacity limited to one, the first step would be to assess if face detection is – or is not – capacity limited to one. To do so, the first step would be to assess the detection of small quantities between 2 – 4 faces using subitizing paradigms. If detection capacity exceeds this small quantity, then ANS paradigms could be adapted to assess the approximate capacity of face detection.

1.2.3 Adapting Numerical Cognition Paradigms for Face Detection

One of the limitations of previous face detection experiments is the use of a single face in the visual environment. While this approach has provided insight into the serial vs parallel nature of the process, it has also led to some discrepancies. The use of a single target face does not adequately test the capacity limit of face detection as detection processing capacities are not overloaded. However, subitizing methods could easily be adapted for investigating multiple face detection. Previous subitizing studies have typically avoided presenting complex objects in their displays to eliminate non-numerical variables that may influence findings. As a standard, dots or circles are used instead of complex visual stimuli to control for effects of visual properties such as luminance, surface area, size, and density (De Marco & Cutini, 2020). However, these dots and circles could be easily replaced with multiple faces and these visual properties hijacked to measure their effect on detection. The face detection literature already points to a different pattern in detecting single faces compared to other single stimuli. Examining subitizing for faces should enrich our understanding of this detection advantage, by establishing whether it extends beyond a single face.

There are already existing similarities between subitizing and face detection paradigms, specifically Purcell and Stewart's (1986; 1988) FDE. Both of these paradigms rely on minimum presentation time to measure the detection of visually presented items. Moreover, the FINSTs theory underlying subitizing is originally a vision theory concerned with how items in the visual environment are located and tracked – processes that have much in common with detection. From a FINSTs perspective, replacing the 1 – 8 dots in subitizing studies with 1 – 8 faces could indicate how faces are individuated and detected by looking at both accuracy and reaction time measures. If the results suggest a pattern similar to subitizing, whereby a limited number of faces are rapidly and accurately enumerated, then this could be taken as evidence supporting parallel processing of more than one face. However, if accuracy decreases and reaction times increase even over the standard subitizing range, this could be taken as evidence for strictly serial processing. Additionally, comparing performance in subitizing upright faces to subitizing inverted faces or other visually similar items could help to understand the tuning of the face detection template, that is, what counts as a face to the visual system. Given that single face detection seems impervious to orientation (i.e. upright or inverted faces), an interesting question will be whether upright and inverted faces load onto the same face detection capacity limits.

Once a foundation is established on the pattern of results for the subitizing of faces, these tasks could be adapted further to be visual search tasks. For instance, multiple upright faces could be embedded among inverted faces or even naturalistic scenes. If multiple upright faces are easily detected from naturalistic settings, then this would imply a parallel process of face detection. Furthermore, if upright faces are detected serially when embedded among inverted faces, then this would replicate Lewis and Edmonds (2005) findings and support a staged detection process.

There is precedent for adapting numerical cognition studies and subitizing studies to investigate how multiple items in the visual environment are detected. For instance, Vuilleumier & Rafal (2000) used subitizing paradigms with simple geometric shapes such as triangles and stars to investigate object detection in patients with visual extinction. They found evidence of parallel processing in patients with right partial lobe damage and left hemifield neglect. When these patients were shown objects in the left hemifield, the detection was impaired as expected. However, when objects were presented in the right hemifield or both hemifields simultaneously, they could accurately enumerate and detect the items displaying similar reaction times to controls. Subitizing has also been used to assess the parsing of biologically relevant stimuli from complex visual backgrounds. Railo et al. (2016) presented displays containing 1 – 6 human targets for 50 ms to assess visual item enumerations (See Figure 1.6 for examples). The faces of these human targets were not clearly visible, but up to three targets could be accurately and rapidly detected, while additional targets incurred a detection cost

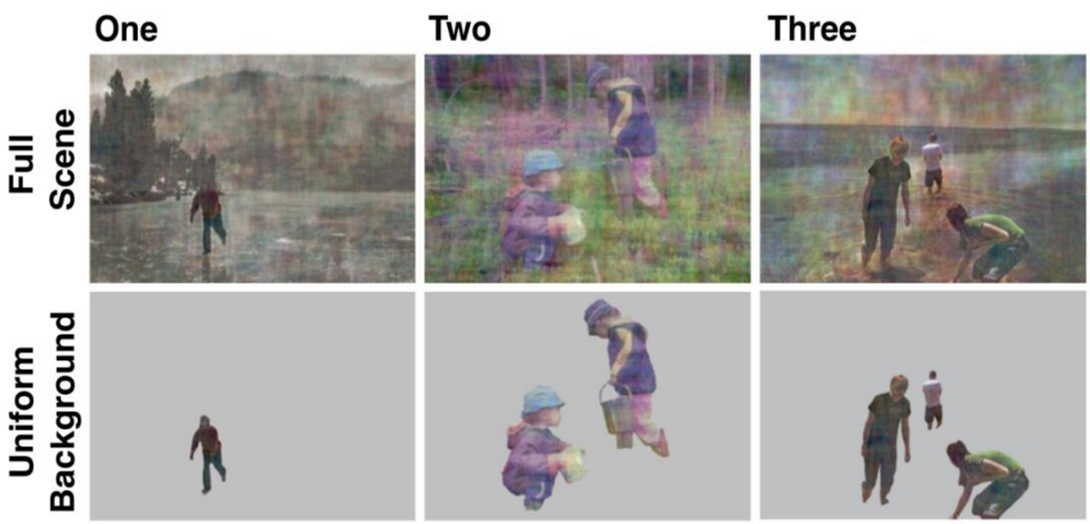

Figure 1.6 Example displays adapted from Railo et al (2016) of One, Two, and Three bodies in full scenes and uniform backgrounds.

Moreover, Puce et al. (2013) demonstrate that studies of face perception can incorporate multiple faces to assess their necessary aims. In an ERP study, the authors found that increases in the number of faces elicited corresponding increases in the amplitude of the N170 – a face-selective measure. However, this task did not involve parsing faces from complex backgrounds as would be expected in detection, and as the authors mention, it also did not involve determining whether the stimulus was a face or not. Similar amplitude responses in the N170 have been found for other non-face objects (Guillaume et al., 2009; Rosburg et al., 2009; Rossion et al., 2003), and so increases in the number of these non-face objects may produce the same results.

Nonetheless, Puce et al. (2013) show that multiple faces can be incorporated to assess face perception. In addition, Vuilleumier & Rafal (2000) and Railo et al. (2016) effectively show that subitizing paradigms can be adapted to investigate aims beyond just numerical cognition. But these studies also highlight an important consideration; paradigms from numerical cognition literature must be adapted appropriately to investigate face detection rather than investigating subitizing alone. Subitizing paradigms usually incorporate simple circular objects on plain backgrounds. A good starting point would be to replace these circles with faces. However, to adequately investigate the nuances of face detection, paradigms should be built further to incorporate inverted faces and non-face objects, as well as complex visual backgrounds and other scene contexts. The next section lays out an overview of the experimental work presented in this thesis to address the capacity limits of face detection.

1.3 Overview of the Current Work

A common characteristic of most face detection literature discussed in this chapter so far is that it neglects to consider multiple face detection. While discoveries have been made regarding the qualitative nature of the detection template and its visual properties, substantially less is known about the quantitative aspect. The experimental chapters of this thesis have three main aims. The first aim is to compare multiple target detection for faces and other types of stimuli. The second aim is to assess capacity limits in face detection and investigate the serial-vs-parallel nature of face template matching. The third aim focuses on multiple face detection testing how it is affected by different viewing conditions, including aspects of the task and presentations

Chapter 2 centres on the first aim of distinguishing between multiple face and multiple non-face detection. A ‘subitizing of faces’ methodology is used to test whether multiple faces can be detected simultaneously and whether their detection outperforms multiple non-faces. Experiment 1 uses an ‘absolute subitizing’ approach, comparing detection accuracy and reaction time for 1 – 8 upright faces, inverted faces, upright non-faces, and inverted non-faces in brief displays. Experiment 2 replicates Experiment 1 but presents displays until response to assess the effect of display exposure duration. Experiment 3 adopts a ‘categorical subitizing’ approach, embedding multiple upright faces amongst non-faces to test whether multiple faces can be separated from other visual objects concurrently.

Chapter 3 assesses the capacity limits of face detection and the serial-*vs*-parallel nature of the template matching process. In this chapter, a new methodology is devised whereby participants saw face and non-face items in ‘fixed’ displays (all one stimulus type) or ‘mixed’ displays (a combination of both types). For each display, the participants’ task was to indicate whether the items were fixed or mixed. Determining that they are fixed requires all items to be categorised. Experiments 4 and 5 compared the detection of faces to Fourier-transformed scrambled faces in displays of two vs three items and two vs four items, respectively. Experiments 6 and 7 replicated Experiments 4 and 5 but compared faces to inverted faces. Manipulating item numerosity and type allows for an estimate of the detection cost of each additional item in the display. If increasing the number of faces incurs a detection cost, this will indicate a serial capacity-limited detection process. If adding faces does not incur any detection costs, then face detection could be described as a parallel process.

The experiments in Chapter 4 assess multiple face detection in real scenes and the visual properties that may affect it. In the first six experiments, participants are presented with displays containing one, two, three, or four items of a fixed type embedded in different visual contexts. Similar to subitizing, the task is to indicate how many items are present as quickly and accurately as possible. In Experiments 8 and 9, display presentation times are manipulated, and detection efficiencies for Upright, Inverted, or Scrambled Faces in real scenes are measured. Experiments 10 and 11 manipulate background meaningfulness and complexity to investigate their effect on multiple face detection. Real scenes are once again used as visual context in Experiments 12 and 13, but item type and face orientation are manipulated to assess their effects on template matching in multiple face detection. Finally, Experiment 14 uses real photographs in a large-scale single-trial design to evaluate spontaneous multiple face detection as it might occur outside of laboratory settings.

To conclude, Chapter 5 discusses the implications of the current experimental work, contextualising it within face detection literature. First, it shall fill in the gaps regarding the quantitative aspect of the process, discuss what multiple face detection can tell us about the qualitative aspect of the process, and discuss the social and evolutionary advantages of detecting multiple faces. Chapter 5 shall also discuss the methodological implications of the current work before discussing some direction for future investigations into multiple face detection.

Chapter 2 – Subitizing Faces

2.1 Introduction

Our ability to derive social information from faces plays a foundational role in our ability to interact with people. Not only do faces attract more of our visual attention (Langton et al., 2008) but even from a few months old, we begin to preferentially tune to them, processing information about gaze, identity, emotion, and even race (Corkum & Moore, 1998; Farroni et al., 2005; Frank et al., 2009; Kelly et al., 2005, 2019; Pascalis et al., 1998; Scaife & Bruner, 1975). Much of this face processing is rapid but strictly capacity limited to one face at a time (Bindemann et al., 2007; Bindemann, Burton, et al., 2005). However, before any face can be processed, it must be detected by matching regions of the visual environment to a stored face template (Lewis & Ellis, 2003; Robertson et al., 2017; Tsao & Livingstone, 2008).

How many faces we can detect remains an open question, but there is a large body of evidence supporting our ability to detect multiple objects at once. Subitizing refers to the rapid and accurate enumeration of up to three or four items in the visual environment (Kaufman & Lord, 1949). Beyond four items, enumeration accuracy and speed begin to decline. It has been proposed that subitizing is underlined by a general non-numerical process of multiple item individuation that operates in a parallel but capacity-limited manner (Mazza & Caramazza, 2015; Piazza et al., 2011; Trick & Pylyshyn, 1994). Subitizing paradigms involve rapidly presenting displays of 1 – 8 non-salient objects to participants to assess visual item enumeration. Accuracy and reaction times are measured and usually follow the pattern shown in Figure 2.1 (Piazza et al., 2011). For up to four items accuracy is high and reaction times are low, however beyond four items subitizing performance breaks down with lower accuracy and slower reaction times. The items used in subitizing paradigms tend to be non-salient stimuli such as simple circles or squares to avoid interference from confounding variables (De Marco & Cutini, 2020). But despite usually being assessed through simple geometric shapes, it appears that subitizing is a domain-general process (for a review, see Katzin et al., 2019).

The non-domain specificity of subitizing further supports that detecting multiple faces is plausible. Moreover, there are at least two reasons to suggest that detecting multiple faces is better than detecting multiple non-faces. First, many have argued that faces are 'special', in the sense that they follow different perceptual principles than non-face objects (Farah et al., 1998; Kanwisher et al., 1997; Tanaka & Farah, 2007; Yin, 1969a). A consistent finding from face detection and face perception literature is a hierarchy for detecting different classes of objects. Converging evidence from eye-tracking and visual search studies show that detecting a face outperforms detecting a face-like object such as pareidolic or illusionary faces, which in turn outperforms the detection of a non-face object (Crouzet et al., 2010; Crouzet & Thorpe, 2011; Keys et al., 2021; Wardle et al., 2020). Upright schematic faces also require shorter presentation times to be detected than jumbled or inverted schematic faces (Purcell & Stewart, 1986, 1988). Moreover, this greater detection performance appears specific to human faces over other mammal, or animal faces (Simpson, Buchin, et al., 2014; Simpson, Maylott, Leonard, et al., 2019). This finding supports a detection mechanism specialised for detecting a face over another item and points to some specificity in the face detection template.

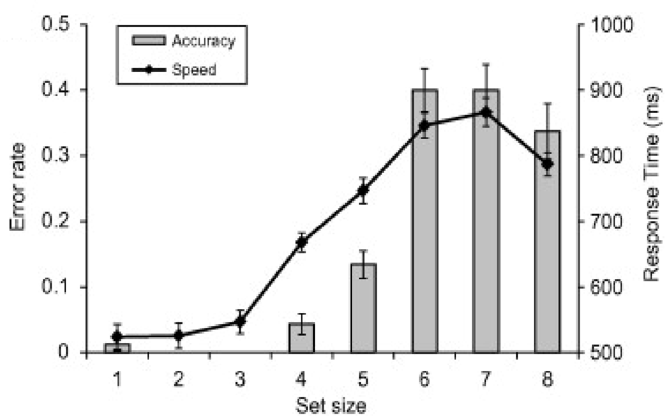

Figure 2.1 Typical subitizing accuracy and reaction times based on set size, adapted from Piazza et al. (2011). Error rates and reactions times are flat over Set Size range 1–3.

However, there is also some evidence to suggest face detection, and therefore the face template is invariant to inversion. Bindemann & Burton (2008) found an inverted face was as likely to capture attention as an upright face. The authors suggest this may be due to some common visual attributes between upright and inverted faces. In the context of face detection, this may emerge in the orientation sensitivity of the detection template. However, the presence of inversion invariance at detection would contrast with inversion effects seen in later face processes (Farah et al., 1995; Yovel & Kanwisher, 2005a). But whether upright and inverted faces are detected in a similar manner requires further investigation. Nonetheless, it appears a detecting a face is biologically and socially prioritised over detecting a face-like or non-face object. In that case, subitizing may even be more efficient for faces than for other items, leading to a higher quantity of faces detected.

The second reason to suspect that multiple faces are detected differently is the presence of a bottleneck strictly constraining later face processes to a limit of one face (Bindemann et al., 2007; Bindemann, Burton, et al., 2005). However, where this bottleneck lies within face perception is unclear. The bottleneck could either originate at the detection stage, constraining detection and subsequent processes. If this was the case, then the subitizing of faces would be limited. But it is also just as likely that the bottleneck is present after the detection stage such that multiple faces can be detected – or subitized – at once. But to assess the capacity limits, suitable paradigms using multiple faces must first be developed.

Except for a few studies (e.g., Bindemann & Burton, 2008), face detection experiments usually utilise a single target within their tasks. Consequently, they have neglected to consider if we can detect multiple faces better than we can detect multiple non-faces. However, there have been successful attempts at incorporating more than one face within detection studies. Evidence from an ERP experiment by Puce et al., (2013) suggests that we might be capable of knowing if more than one face is present. The N170, a face-selective marker, appears to be modulated as the number of faces presented increases. But the task did not compare the detection of multiple faces to the detection of multiple non-faces. Therefore, it is unclear if the greater detection performance towards a single face can extend to multiple faces.

Typical subitizing paradigms present 1 – 8 non-salient objects for very brief durations and assess accuracy and reaction times. The studies in this experimental chapter adapt the basic set up of subitizing paradigms to investigate multiple face detection compared to multiple non-face detection. Experiments 1 and 2 manipulate display exposure duration and present *Upright Faces*, *Inverted Faces*, *Upright Non-Faces*, and *Inverted Non-Faces* in Set Sizes of One to Eight. Each display only contains one type of item and is presented for 16 ms in Experiment 1 or until the participant responds in Experiment 2. Participants are tasked with indicating through button press how many items they saw in the display as quickly and accurately as possible. Suppose multiple face detection is a more rapid and accurate process than multiple non-face detection. In that case, this should emerge as greater performance for face items within the subitizing range of up to four items. Moreover, if the face detection template was orientation invariant, then *Upright* and *Inverted Faces* detection are expected to be detected similarly. Experiment 3 adopts a categorical detection approach, embedding faces and non-faces in different combinations within a single display of four items. Face detection is still expected to outperform non-face detection under these conditions. Furthermore, if the detection template is orientation invariant, this should be evident in poor detection efficiency for *Upright* and *Inverted Faces* within the same display.

2.2 Experiment 1: Absolute Subitizing: 16 ms Exposure Time

This first experiment investigates if faces are detected faster and with greater accuracy than non-faces. It directly compares *Upright Faces*, *Inverted Faces*, *Upright Non-Faces*, and *Inverted Non-Faces* in Set Sizes of One, Two, Three, Four, Five, Six, Seven, and Eight items. Participants are presented with a single type of item in each display and given the task to respond with how many items they saw as quickly and accurately as possible. Crucially, each display was only presented for 16 ms (monitor frame refresh rate). The 16 ms exposure duration was adopted from prior subitizing studies as a minimum presentation threshold from which participants could accurately enumerate items on screen (Inglis & Gilmore, 2013). This limited exposure duration also ensures that face processing is constricted to the detection stage alone. As with prior subitizing experiments, accuracy and reaction time will be measured as an indicator of detection performance (Piazza et al., 2011). If a specialised detection mechanism is present for faces, then face stimuli should be detected faster and more accurately than non-face stimuli. Moreover, this should be evident within the subitising range consisting of the smaller Set Sizes of One through Four.

2.2.1 Methods

2.2.1.1 Participants

Seventy participants were recruited from the University of York's Human Participant Pool (SONA) and completed the experiment in exchange for a small payment. A total of thirty participants were excluded from the final analysis; 24 due to 0% accuracy in at least one condition, 3 due to accuracy scores below 70% in the training block, and 3 due to technical issues during testing. The final sample (N = 40) consisted of 36 females and 4 males (age range 18–24; M = 19.80, SD = 1.40).

2.2.1.2 Design and Stimuli

A local bank of 288 faces was created by collecting face images from AI Generated Faces (Karras et al. & Nvidia, 2018), MR2 face bank (Strohminger et al., 2016) and other online sources. The bank contained an equal distribution of faces divided into 12 categories based on sex (male and female), age (young and old adults), and ethnicity (Asian, Black, and Caucasian; see Prunty et al., (2022) for details of demographic categorisation). Each face was segmented from the background and cropped to outline using the InterFace software package (Kramer et al., 2017). The resulting face image was resized to 380 pixels wide × 570 pixels high to create *Upright Face* stimuli. *Inverted Face* stimuli were created by rotating *Upright Faces* 180° in the picture plane.

Boots were chosen as the non-face item category as they can be divided into 12 distinguishable categories and retain the same distribution in item variability as the face categories. A total of 288 boot images were collected from online sources, with an equal distribution based on sex (male and female), left- or right-facing, and type (Rain, Snow, Cowboy). To create the *Upright Non-Face* stimuli, boot images were cropped to outline and resized to 380 pixels wide × 570 pixels high. The *Upright Non-Face* stimuli were rotated 180° in the picture plane, creating *Inverted Non-Face* stimuli.

Upright Face displays were generated by embedding between 1 and 8 faces on a blank grey background. Each display contained 25 predetermined locations in a hexagonal grid pattern within the central 75% of the screen where the face images could be allocated. For Set Size One, a single random location was chosen from the predetermined set. For Set Sizes greater than One, an item location-allocation method was developed. To avoid groupitizing, whereby the clustering of items in small subsets facilitates subitizing, equal distances were maintained between faces (Anobile et al., 2020; Starkey & McCandliss, 2014). First, a face is allocated to a randomly selected starting point from the predetermined locations. Due to the hexagonal grid pattern, the starting location contains three unoccupied adjacent locations where the second face could be allocated. Consequently, the second location could only contain up to two unoccupied adjacent locations. Faces can then continue to be assigned to the following unoccupied neighbouring location in the same manner until the number of faces meets that of the required Set Size. The displays were created live on each iteration of the experiment, so a new random set of faces was used for each display. However, no upright face item was repeated within the same display or throughout the experiment. The same procedure was used to generate *Inverted Face*, *Upright Non-Faces*, and *Inverted Non-Face* displays. Example displays are presented in Figure 2.2.

The experiment was created and hosted using MATLAB (R2019a) with the Psychtoolbox extension (v 3.0.16) (Brainard, 1997; Kleiner et al., 2007). The study was run on an HP EliteDesk 800 G3 TWR with an i7 intel core, and the displays were presented on a 23-inch HP EliteDisplay E223 monitor with a 1920 × 1080 resolution and 60Hz.

The within-subjects factors of Set Size (*One, Two*, *Three, Four, Five, Six, Seven, Eight*) and Item Type (*Upright Face, Inverted Face, Upright Non-Face, Inverted Non-Face*) were manipulated in a fully counterbalanced 8 × 4 factorial design.

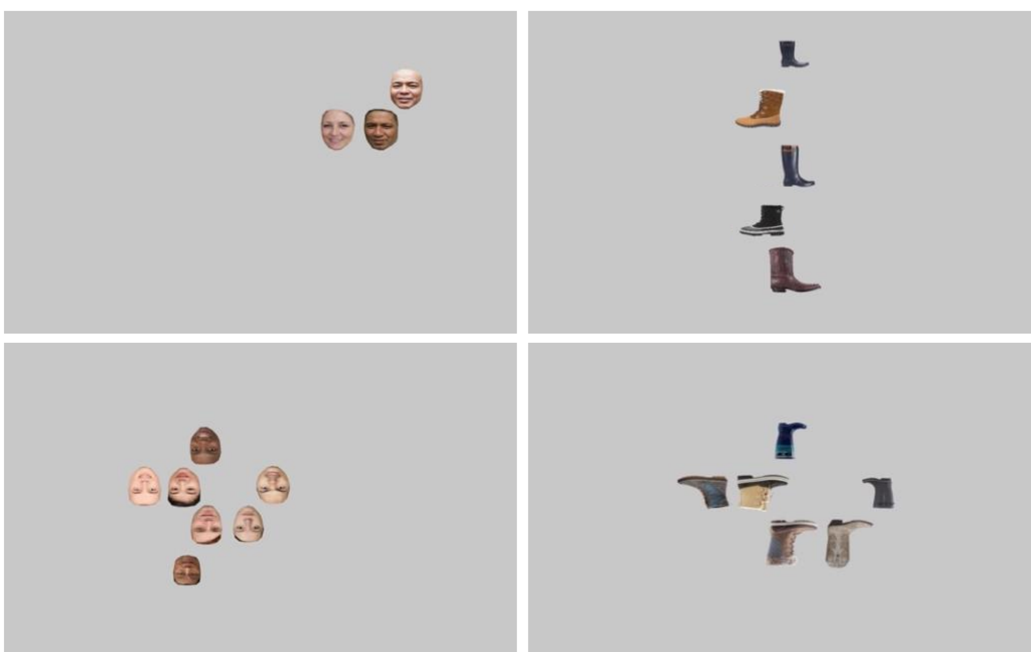

Figure 2.2 Example displays used in Experiments 1 and 2. Top left – Three Upright Faces; top right – Five Upright Non-Faces; bottom left – Seven Inverted Faces; bottom right – Six Inverted Non-Faces.

2.2.1.3 Procedure

Participants sat in front of the screen and asked to adjust themselves to be at a comfortable distance such that all sides of a red rectangle presented on the screen were visible. This was to ensure that all items on the screen were viewable to the participants. The participants were tasked with indicating how many items they saw in each display as quickly and accurately as possible. The response set-up involved the participants placing the four fingers of the left hand on the ‘1’, ‘2’, ‘3’, and ‘4’ keyboard number keys, and the four fingers of the right hand on the ‘5’, ‘6’, ‘7’, and ‘8’ keyboard number keys. To respond, participants were instructed to press the number key that corresponded to how many items they thought were on the screen. First, the participants were given a training block of 16 trials (2 trials per response key) to familiarise them with the response set-up. Training trials consisted of a fixation cross presented for 500 ms followed by a number between 1 – 8 for 16 ms and then a visual mask with a response prompt screen.

At the end of the training block, the experimental blocks were presented. Each experimental trial began with a 500 ms fixation cross, a 16 ms display exposure time, and then a response prompt screen containing a visual mask. The experiment consisted of 12 experimental blocks presented in random order. Each block consisted of 64 trials of a single condition, and there were three blocks for each of the *Upright Faces*, *Inverted Faces*, *Upright Non-Faces*, and *Inverted Non-Faces* conditions. Consequently, 192 trials contributed to each condition, with 24 trials per set size. Participants were given the opportunity to take short breaks between the blocks. The entire experiment took approximately 40 min to complete.

2.2.2 Results and Discussion

Mean accuracy, reaction time, and full two-way ANOVA of Set Size and Item Type for all conditions are reported as supplementary material in the appendices. Simple main effects found no accuracy differences based on orientation for *Face* and

Non-Face stimuli at any Set Size. Further t-tests comparing *Upright Faces* (M = 62.64%, SE = 1.56%), and *Inverted Faces* conditions (M = 63.19%, SE =1.56%), ($t(319) = -0.81$, $p = .421$) and *Upright Non-Faces* (M = 61.17%, SE = 1.42%) and *Inverted Non-Faces* (M = 61.13%, SE = 1.44%), ($t(319) = 0.49$, $p = .962$) revealed no significant differences as well. The data was averaged across orientations to create new combined *Face* and *Non-Face conditions*. Figure 2.3 (A) displays the new combined data for accuracy and reaction time for Experiment 1.

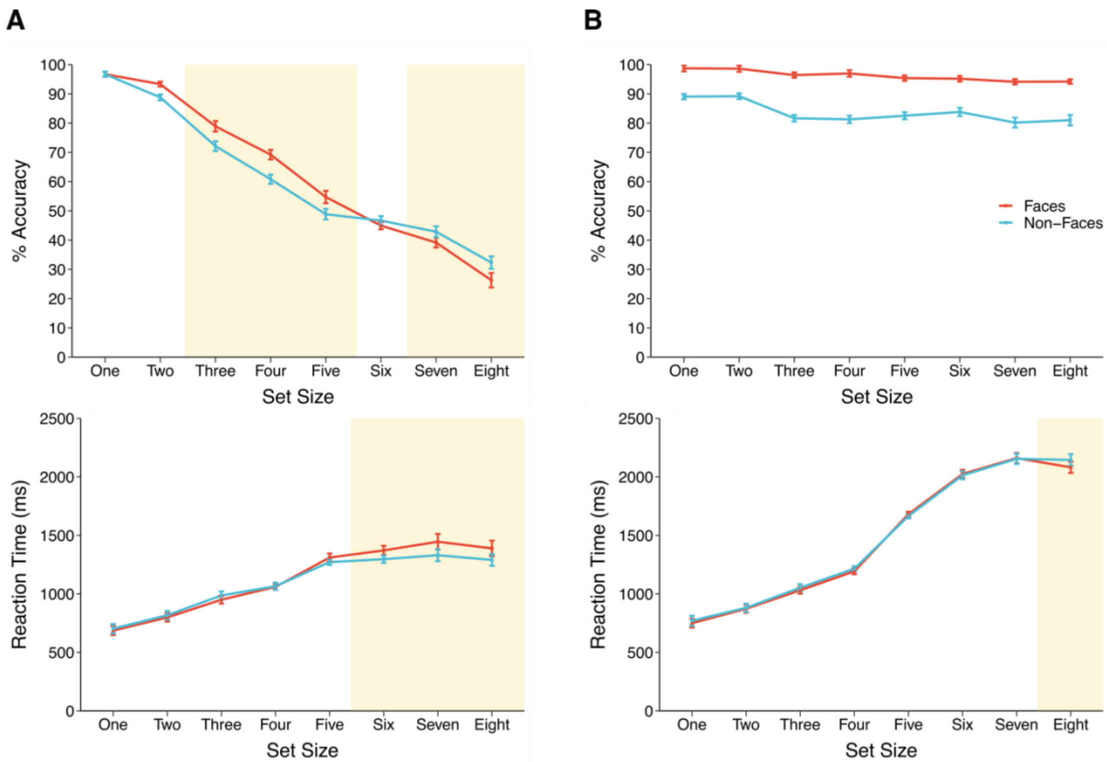

Figure 2.3 Mean percentage accuracy (%) and reaction time results (ms) for Face and Non-Face conditions in Experiment 1 (A) and Experiment 2 (B). Error bars show within-subject standard error (Cousineau, 2005). Yellow shaded areas indicate a significant difference between Faces and Non-Faces at a particular set size.

For concision of the results section, each individual condition will be referred to by the Set Size numeral followed by the Item Type initial. For instance, *Faces* at Set Size One will be 1F, while *Non-Faces* at Set Size Two will be 2N, and so forth.

2.2.2.1

2.2.2.1 *Detection Accuracy*

To investigate detection accuracy for faces compared to non-faces, a two-way ANOVA with repeated measures of Set Size (*One, Two*, *Three, Four, Five, Six, Seven, Eight*) and Item Type (*Faces*, *Non-Faces*) was conducted on the new combined accuracy data. The analysis revealed a significant main effect of Set Size, with accuracy decreasing as the Set Size increased [F (7, 273) = 248.50, $p < .001$, $\eta^2 = 0.86$; One, M = 96.72%, SE = 0.88%; Two, M = 91.07%, SE = 0.93%; Three, M = 75.52%, SE = 1.73%; Four, M = 65.03%, SE = 1.60%; Five, M = 51.8%, SE = 1.96%; Six, M = 45.81%, SE = 1.39%; Seven, M = 41.02%, SE = 1.77%; Eight, M = 29.32%, SE = 2.28%].As well as a main effect of Item Type, with greater overall accuracy for *Faces* (M = 62.92%, SE = 1.58%) compared to *Non-Faces* (M = 61.15%, SE = 1.56%), [F (1, 39) = 9.00, $p = .005$, $\eta^2 = 0.19$]. A significant interaction effect between Set Size and Item Type was also found [F (7, 273) = 12.04, $p < .001$, $\eta^2 = 0.24$].

Simple main effects revealed significant differences in accuracy for *Faces* as each Set Size increased [F (1, 564) = 224.91, $p < .001$, $\eta^2 = 0.74$]. Detection accuracy for 1F and 2F is similarly high but decreased significantly with each additional face, and no differences were found between 6F and 7F either, (1F, M = 96.72%, SE = 0.85%; 2F, M = 93.33%, SE = 0.85%; 3F, M = 78.91%, SE = 1.79%; 4F, M = 69.22%, SE = 1.63%; 5F, M = 54.74%, SE = 2.11%; 6F, M = 44.95%, SE = 1.25%; 7F, M = 39.17%, SE = 1.68%; 8F, M = 26.3%, SE = 2.46%).

Accuracy for *Non-Faces* also decreased as Set Size increased [F (1, 564) = 177.491, $p < .001$, $\eta^2 = 0.69$], with significant differences found across all Set Sizes except between 5N, 6N, and 7N, (1N, M = 96.72%, SE = 0.91%; 2N, M = 88.8%, SE = 1.01%; 3N, M = 72.14%, SE = 1.67%; 4N, M = 60.83%, SE = 1.56%; 5N, M = 48.85%, SE = 1.81%; 6N, M = 46.67%, SE = 1.54%; 7N, M = 42.87%, SE = 1.86%; 8N, M = 32.34%, SE = 2.1%).

Simple main effects also revealed no difference in detection accuracy at Set Size One between 1F and 1N. However, *Faces* were detected more accurately than *Non-Faces* at Set Sizes Two to Five. At Set Size Six, detection accuracy was not significantly different between 6F and 6N, but beyond that, a switch over occurs where 7N and 8N were detected more accurately than 7F and 8F respectively, [Set Size One, $F(1, 312) = 0.00$, $p = .999$, $\eta^2 = 0.00$; Set Size Two, $F(1, 312) = 8.49$, $p = .004$, $\eta^2 = 0.03$; Set Size Three, $F(1, 312) = 18.97$, $p < .001$, $\eta^2 = 0.06$; Set Size Four, $F(1, 312) = 29.09$, $p < .001$, $\eta^2 = 0.09$; Set Size Five , $F(1, 312) = 14.33$, $p < .001$, $\eta^2 = 0.04$; Set Size Six, $F(1, 312) = 1.22$, $p = .270$, $\eta^2 = 0.00$; Set Size Seven, $F(1, 312) = 5.66$, $p = .018$, $\eta^2 = 0.02$; Set Size Eight, $F(1, 312) = 15.1$, $p < .001$, $\eta^2 = 0.05$].

2.2.2.2 Detection Reaction Time

To further assess detection reaction time for faces compared to non-faces, another two-way repeated measures ANOVA with the same repeated measures was conducted on the combined RT data. The analysis found a significant main effect of Set Size, with longer RTs as Set Size increased [$F(7, 273) = 44.87$, $p < .001$, $\eta^2 = 0.5$; One, M = 694 ms, SE = 39 ms; Two, M = 808 ms, SE = 36 ms; Three, M = 968 ms, SE = 34 ms; Four, M = 1063 ms, SE = 28 ms; Five, M = 1291 ms, SE = 30 ms; Six, M = 1334 ms, SE = 35 ms; Seven, M = 1388 ms, SE = 59 ms; Eight, M = 1340 ms, SE = 58 ms]. No main effect of Item type was found with similar overall RTs for *Faces* (M = 1127 ms, SE = 42 ms) and *Non-Faces* (M = 1095 ms, SE = 37 ms), [$F(1, 39) = 3.60$, $p = .065$, $\eta^2 = 0.08$]. However a significant interaction effect between Set Size and Item Type was found [$F(7, 273) = 2.84$, $p < .001$, $\eta^2 = 0.07$].

Simple main effects revealed significant differences in RTs for *Faces* as Set Size increased. Overall, RTs increased as the number of faces to detect increased, [$F(7, 564) = 45.35$, $p < .001$, $\eta^2 = 0.37$]. For smaller Set Sizes between 1F and 4F, RTs for each Set Size were not significantly different from their immediate neighbouring Set Size, (IF, M = 686 ms, SE = 39 ms; 2F, M = 799 ms, SE = 36 ms; 3F, M = 950 ms, SE = 33 ms; 4F, M = 1060 ms, SE = 25 ms). However, RTs for the smaller Set Sizes were overall significantly quicker than RTs for larger Set Sizes of 5F to 8F, and no differences were found within the larger Set Sizes, (5F, M = 1311 ms, SE = 34 ms; 6F, M = 1372 ms, SE = 39 ms; 7F, M = 1446 ms, SE = 67 ms; 8F, M = 1389 ms, SE = 66 ms).

Reaction times for *Non-Faces* followed the same pattern as *Faces*, [F (7, 564) = 30.85, $p < .001$, $\eta^2 = 0.28$]. Again, for between 1N and 4N, RTs for each Set Size were not significantly different from their immediate neighbouring Set Size, (1N, M = 703 ms, SE = 40 ms; 2N, M = 818 ms, SE = 36 ms; 3N, M = 986 ms, SE = 35 ms; 4F, M = 1066 ms, SE = 30 ms). Smaller Set Sizes were also found to be significantly quicker overall than larger Set Sizes of 5N to 8N, and no differences were found within these larger Set Sizes again, (5N, M = 1272 ms, SE = 25 ms; 6N, M = 1297 ms, SE = 31 ms; 7N, M = 1331 ms, SE = 51 ms; 8N, M = 1290 ms, SE = 50 ms).

Simple main effects also revealed no difference in RT between detecting a *Face* and a *Non-Face* at Set Size One through Five. Beyond that, *Non-Faces* were detected quicker than *Faces* at Set Sizes Six, Seven, and Eight [Set Size One, F (1, 312) = 0.20, $p = .656$, $\eta^2 = 0.00$; Set Size Two, F (1, 312) = 0.26, $p = .613$, $\eta^2 = 0.01$; Set Size Three, F (1, 312) = 0.97, $p = .326$, $\eta^2 = 0.02$; Set Size Four, F (1, 312) = 0.02, $p = .876$, $\eta^2 = 0.03$; Set Size Five, F (1, 312) = 1.11, $p = .292$, $\eta^2 = 0.04$; Set Size Six, F (1, 312) = 4.20, $p = .041$, $\eta^2 = 0.01$; Set Size Seven, F (1, 312) = 9.89, $p = .002$, $\eta^2 = 0.03$; Set Size Eight, F (1, 312) = 7.31, $p = .007$, $\eta^2 = 0.02$].

Initial analyses in Experiment 1 found no differences in orientation for either Faces or Non-Faces across Set Sizes. This could point to an orientation-invariant element to face detection that contrasts with inversion effects reported in later stages of face perception. Further analyses directly compared the detection accuracy and RTs of *Faces* to *Non-Faces* in Set Sizes of up to eight items. Analyses show that while detection times are similar, *Faces* are detected more accurately than *Non-Faces* at smaller Set Sizes (up to Four or Five). Within this range of faces, accuracy decreased incrementally to ~50% at Set Size five, but detection times were similar. At larger Set Sizes, a switchover between *Faces* and *Non-Faces* occurs where the times taken to detect *Faces* are longer, and accuracy is poorer. But face detection accuracy is still above chance even at Set Size Eight, suggesting that up to eight faces can still be detected but with great difficulty. These results support a possible detection mechanism preferential to *Faces* over *Non-Faces* (Purcell & Stewart, 1986, 1988). However, they also point to a limit to this advantage, which is constrained to smaller numbers of faces, e.g. up to four.

The displays in Experiment 1 were presented for a *very* brief time of 16 ms. This rapid display exposure time was useful in constraining perception strictly to the detection stage, and in some conditions, participants were over 95% accurate in their responses. However, a short exposure time also led to the exclusion of a substantial number of participants before analysis due to scoring 0% in conditions at the larger Set Sizes. In the next experiment, we increase the display exposure time until response and increase the total number of participants recruited.

2.3 Experiment 2: Absolute Subitizing: Until Response

Experiment 2 extends upon Experiment 1 to investigate if face detection outperforms non-face detection. Here all aspects of Experiment 1 are replicated, except exposure duration time limits are removed. Allowing participants as much time as needed to respond should increase accuracy across the entire experiment. However, faces should still be detected faster at smaller Set Sizes of One through Four compared to larger Set Sizes of Five through Eight.

2.3.1 Methods

2.3.1.1 Participants

Sixty-one participants were recruited from the University of York's Human Participant Pool (SONA) and completed the experiment in exchange for a small payment. One participant was excluded from the final analysis due to accuracy scores below 70% in the training block. The final sample (N = 60) consisted of 57 females and 3 males (age range 18–59; M = 20.58, SD = 5.42).

2.3.1.2 Design and Stimuli

The design and stimuli for Experiment 2 were identical to that of Experiment 1.

2.3.1.3 Procedure

The procedure, task, and experimental set-up were identical to Experiment 1, except displays were presented until participants responded.

2.3.2 Results and Discussion

Mean accuracy, reaction time, and full two-way ANOVA of Set Size and Item Type for all conditions are reported in as supplementary material in the appendices. As with Experiment 1, simple main effects found no accuracy differences based on orientation for *Face* and *Non-Face* stimuli at any Set Size. Further t-tests comparing *Upright Face* (M = 90.09%, SE = 0.69%) and *Inverted Face* conditions (M = 89.44%, SE = 0.70%), ($t(479) = 1.69$, $p = .091$) showed no significant difference either. However, a small but significant difference was found for *Upright Non-Faces* (M = 90.44%, SE = 0.69%) over *Inverted Non-Faces* (M = 89.63%, SE = 0.72%), ($t(479) = 1.69$, $p = .091$). Nonetheless, accuracy was at ceiling overall, so data were combined into new *Face* and *Non-Face* conditions. The new combined data accuracy and RT data for Experiment 2 are displayed in Figure 2.3 (B). The same naming convention used for the results section in Experiment 1 is used here.

2.3.2.1 Detection Accuracy

To investigate detection accuracy for faces compared to non-faces, a two-way ANOVA with repeated measures of Set Size (*One, Two*, *Three, Four, Five, Six, Seven, Eight*) and Item Type (*Faces*, *Non-Faces*) was conducted on the new combined accuracy data. The analysis found a significant main effect of Set Size with accuracy decreasing slightly as Set Size increased, [$F(7, 413) = 39.05$, $p < .001$, $\eta^2 = 0.40$, One, M = 98.68%, SE = 1.08%; Two, M = 96.7%, SE = 1.03%; Three, M = 95.28%, SE = 0.94%; Four, M = 94.17%, SE = 0.85%; Five, M = 89.13%, SE = 0.95%; Six, M = 81.48%, SE = 1.21%; Seven, M = 83.18%, SE = 1.32%; Eight, M = 80.59%, SE = 1.74%]. No main effect of Item Type was found between *Faces* (M = 89.77%, SE = 1.10%) compared to *Non-Faces* (M = 90.01%, SE = 1.18%), [$F(1, 50) = 0.75$, $p = .391$, $\eta^2 = 0.01$], and no significant interaction effect was found between Set Size and Item Type [$F(7, 413) = 0.49$, $p = .839$, $\eta^2 = 0.01$].

Overall detection accuracy for *Faces* and *Non-Faces* was high at all Set Sizes, IF, M = 98.75%, SE = 1.06%; 2F, M = 96.42%, SE = 0.92%; 3F, M = 95.38%, SE = 0.93%; 4F, M = 94.13%, SE = 0.92%; 5F, M = 89.06%, SE = 0.91%; 6F, M = 81.67%, SE = 1.13%; 7F, M = 82.53%, SE = 1.21%; 8F, M = 80.17%, SE = 1.68%; 1N, M = 98.61%, SE = 1.09%; 2N, M = 96.98%, SE = 1.13%; 3N, M = 95.17%, SE = 0.95%; 4N, M = 94.2%, SE = 0.78%; 5N, M = 89.2%, SE = 0.99%; 6N, M = 81.28%, SE = 1.28%; 7N, M = 83.82%, SE = 1.42%; 8N, M = 81.01%, SE = 1.8%).

2.3.2.2 Detection Reaction Time

To further assess detection reaction time for faces compared to non-faces, another two-way repeated measures ANOVA with the same repeated measures was conducted on the combined RT data. The analysis revealed a significant main effect of Set Size, with longer RTs as Set Size increased [F (7, 413) = 242.10, $p < .001$, $\eta^2 = 0.80$; One, M = 760 ms, SE = 41 ms; Two, M = 877 ms, SE = 35 ms; Three, M = 1042 ms, SE = 30 ms; Four, M = 1204 ms, SE = 23 ms; Five, M = 1673 ms, SE = 20 ms; Six, M = 2019 ms, SE = 34 ms; Seven, M = 2157 ms, SE = 44 ms; Eight, M = 2113 ms, SE = 49 ms]. No main effect of Item type was found with similar overall RTs for *Faces* (M = 1474 ms, SE = 34 ms) and *Non-Faces* (M = 1487 ms, SE = 35 ms), [F (1, 59) = 2.88, $p = .095$, $\eta^2 = 0.05$]. However a significant interaction effect between Set Size and Item Type was found [F (7, 413) = 2.76, $p = .008$, $\eta^2 = 0.04$].

Simple main effects revealed significant differences in RTs for *Faces* as Set Size increased. Overall the RT pattern was similar to that of Experiment 1. Reaction times increased as the number of faces to be detected increased, [F (7, 826) = 233.09, $p < .001$, $\eta^2 = 0.66$]. Within the smaller Set Sizes of 1F to 4F, no significant differences in RT were found between immediate neighbouring Set Sizes, (1F, M = 750 ms, SE = 38 ms; 2F, M = 873 ms, SE = 34 ms; 3F, M = 1031 ms, SE = 29 ms; 4F, M = 1193 ms, SE = 25 ms). Moreover, RTs for the smaller Set Sizes were overall significantly quicker than RTs for larger Set Sizes of 5F to 8F, and no differences were found between these larger Set Sizes (5F, M = 1681 ms, SE = 21 ms; 6F, M = 2025 ms, SE = 35 ms; 7F, M = 2160 ms, SE = 44 ms; 8F, M = 2082 ms, SE = 48 ms).

The time taken to detect *Non-Faces* also increased with Set Size. Reaction times for 1N (M = 770 ms, SE = 43 ms) and 2N (M = 881 ms, SE = 35 ms) were quickest overall and not significantly different from each other. The 3N (M = 1053 ms, SE = 30 ms) and 4N (M = 1215 ms, SE = 22 ms) were also not significantly different from each other and detected faster than larger Set Sizes. Reaction times for 5N (M = 1665 ms, SE = 19 ms) differed significantly from all other Set Sizes, but no differences in RTs were found between the remaining *Non-Face* Set Sizes (6N, M = 2013 ms, SE = 33 ms; 7N, M = 2154 ms, SE = 44 ms; 8N, M = 2145 ms, SE = 50 ms).

Simple main effects also revealed no difference in RTs between detecting a *Face* and a *Non-Face* except at Set Size Eight, where detecting 8N took significantly longer than 8F [Set Size One, F (1, 472) = 1.52, p = .218, η^2 = 0.00; Set Size Two, F (1, 472) = 0.27, p = .605, η^2 = 0.00; Set Size Three, F (1, 472) = 1.82, p = .177, η^2 = 0.00; Set Size Four, F (1, 472) = 1.84, p = .175, η^2 = 0.00; Set Size Five, F (1, 472) = 1.00, p = .319, η^2 = 0.00; Set Size Six, F (1, 472) = 0.54, p = .462, η^2 = 0.00; Set Size Seven, F (1, 472) = 0.14, p = .713, η^2 = 0.00; Set Size Eight, F (1, 472) = 15.13, p < .001, η^2 = 0.03].

Experiment 2 compared *Face* and *Non-Face* detection without display exposure time limits. As with Experiment 1, no differences were found between the two face orientations or the two non-face orientations. When participants had as much time as needed to respond, high accuracy was seen across all Set Sizes, and no differences were found between *Faces* and *Non-Faces*. While no effects are seen in face detection accuracy, there appears to be relatively similar high accuracy at smaller Set Sizes (up to four faces) followed by a steady but non-significant decline at larger Set Sizes (five or more faces). Reaction time results clearly highlight the distinction between smaller and larger Set Sizes. A boundary can be drawn at *Five Faces* where smaller Set Sizes are detected in a similar fast manner and larger Set Sizes in a similar but slower manner. This pattern of results echoes that of Experiment 1, where face detection for up to four faces was seen to be quicker and more accurate.

2.4 Experiment 3: Categorical Subitizing

The previous experiments supported a faster and more accurate detection mechanism for up to four faces compared to non-faces. Experiments 1 and 2 took an 'absolute detection' approach by embedding a single Item Type in a display and instructing participants to report how many items they saw. This absolute detection approach helped assess if a difference in detection exists between faces and non-faces. Absolute subitizing does not require the observer to discern between the categories of items displayed to them. However face detection does require faces to be parsed from other items, consequently a slightly modified subitizing paradigm is required to assess this.

Experiment 3 adopts a 'categorical detection' approach, embedding faces and non-faces in a single display to assess if face detection outperforms non-face detection under these circumstances. It directly compares different upright, inverted, and non-face TARGET–distractor combinations (*UPRIGHT–inverted*; *UPRIGHT–non-face*; *INVERTED–upright*; *INVERTED–non-face*; *NON-FACE–upright*; *NON-FACE–inverted)* at Set Sizes of One, Two, Three, and Four targets. Participants were asked to decide as quickly and accurately as possible how many items in the displays presented to them belonged to a specified target category. Each stimulus type could serve as a target in one condition and as a distractor in another condition as specified by the target–distractor conditions above. If face detection is an accurate and rapid process for up to four faces, then efficiency measures combining accuracy and reaction time should reveal no efficiency costs as more faces are added to the display.

2.4.1 Methods

2.4.1.1 Participants

Fifty-one participants were recruited from the University of York's Human Participant Pool (SONA) and completed the experiment in exchange for a small payment. Eleven participants were excluded from the final analysis; 10 due to 0% accuracy in at least one condition and one due to accuracy scores below 70% in the training block. The final sample (N = 40) consisted of 32 females and 8 males (age range 18–33; M = 20.93, SD = 3.76).

2.4.1.2 Design and Stimuli

The same local face bank of 288 faces was used to create *Upright Face* and *Inverted Face* stimuli. After removing the extraneous background, each face image was cropped to 380 pixels wide × 570 pixels high rectangle. *Upright Face* stimuli were rotated 180° in the picture plane to create *Inverted Face* stimuli.

This experiment contained a single *Non-Face* item condition consisting of scrambled faces. *Non-Faces* were created by applying a Fourier phase transformation to the rectangular cropped faces. This transformation randomly scrambles the phase of component spatial frequencies while maintaining overall brightness, contrast, and other lower-level visual properties.

Each display always contained four items in a fixed square formation within the central 75% of the screen. Based on the target Set Size as determined by the condition, the ratio of target and distractor items was varied while counterbalancing for item location. Between 0 – 4 randomly selected targets could be present within a display, with distractors occupying the remaining locations (if available). Figure 2.4 presents example displays. A new set of displays was created for each condition so that no display was used more than once throughout the experiment.

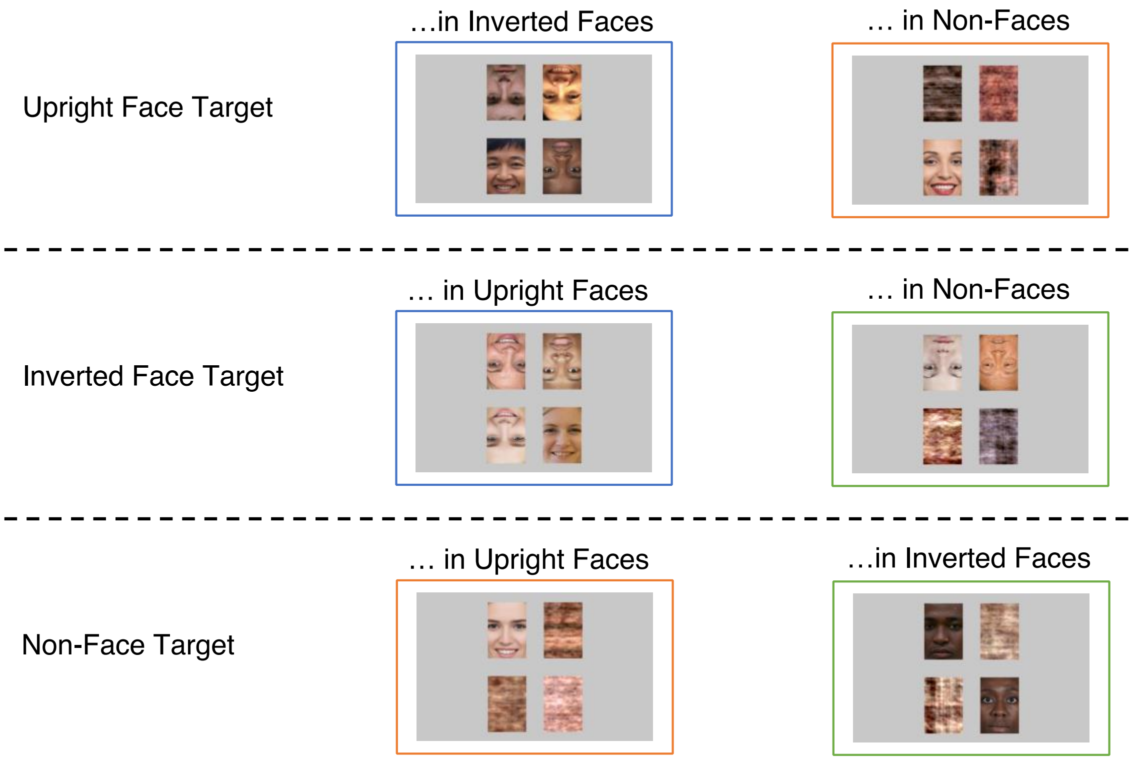

Figure 2.3 Example displays used in Experiment 3. Examples within the same row are grouped by target type. Examples within boxes of the same colour are grouped based on display qualities of item type and ratio.

The experiment was created and hosted using MATLAB (R2019a) with the Psychtoolbox extension (v 3.0.16) (Brainard, 1997; Kleiner et al., 2007). The study was run on an HP EliteDesk 800 G3 TWR with an i7 intel core, and the displays were presented on a 23-inch HP EliteDisplay E223 monitor with a 1920 × 1080 resolution and 60Hz.

The within-subjects factors of Set Size (*Zero*, *One, Two*, *Three, Four*) and TARGET–distractor Type (*UPRIGHT–inverted*; *UPRIGHT–non-face*; *INVERTED–upright*; *INVERTED–non-face*; *NON-FACE–upright*; *NON-FACE–inverted*) and were manipulated in a fully counterbalanced 5 × 6 factorial design.

2.4.1.3 *Procedure*

As with the previous experiments, participants sat in front of the screen and were asked to adjust themselves to be at a comfortable distance such that all sides of a red rectangle presented on the screen were visible to them. They were tasked with indicating as quickly and accurately as possible how many of a specific target they saw in each display. The response set-up involved the participants placing the four fingers of the left hand on the '1', '2', '3', and '4' keyboard number keys, their right-hand index finger on the '0'. To respond, participants were instructed to press the number key that corresponded to how many items they thought were on the screen. First, the participants were given a training block of 25 trials (5 trials per response key) to familiarise them with the response set-up. Training trials consisted of a fixation cross presented for 500 ms, a display containing a number between 0 – 4 for 100 ms, and then a visual mask with a response prompt screen. The experimental blocks were then presented.

Three experimental blocks of 240 trials each were presented to the participants in random order. Each block was organised around a single target item, such that it contained two conditions with one common target and two distractor types, e.g., *Upright Face in Inverted Faces* (1) or *Non-Faces* (2). Before each block, participants were reminded of the task and given the specific target for that block. Twenty-four trials contributed to each condition × Set Size level resulting in 120 trials per condition. The experimental trials began with a 500 ms fixation cross, followed by displays for 100 ms, and then a visual mask with a response prompt screen. Participants were allowed to take short breaks between the blocks, and the entire experiment took approximately 40 minutes to complete.

2.4.2 *Results and Discussion*

Overall accuracy across conditions was 75%, suggesting participants could perform the task. Trials with a reaction time below 15 ms and above 3000 ms (0.68%)

were excluded from the final analysis. Accuracy and reaction time data were used to calculate a new combined Inverse Efficiency Score metric (IES). The IES summarises speed-accuracy performance in a single efficiency measure reported in ms. Higher IES values indicate poorer performance (Bruyer & Brysbaert, 2011; Townsend & Ashby, 1978). Separate analyses of accuracy and reaction time measures are provided as Supplementary Materials in the appendices and support the same conclusions. Figure 2.5 summarises IES data for each condition in Experiment 3.

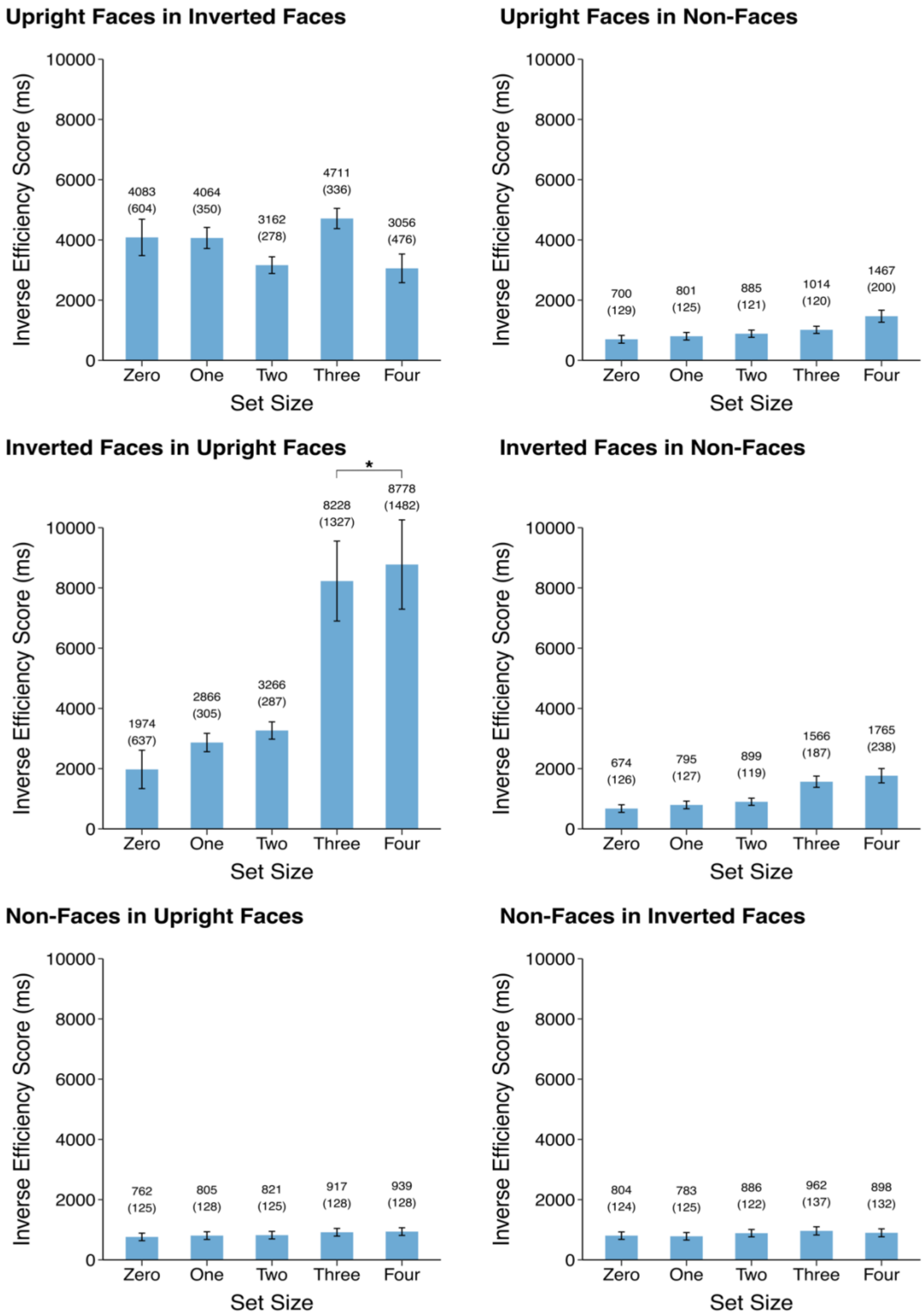

Figure 2.4 Mean IES data for each condition in Experiment 3. Lower scores indicate better efficiency. Error bars show within-subjects standard error (Cousineau, 2005). Bar labels display individual condition means and (SE). Brackets indicate statistical significance between set sizes.

2.4.2.1 Detection Efficiency

To investigate detection efficiency, IES data were subjected to a two-way ANOVA with repeated measures of Set Size (*Zero*, *One, Two*, *Three, Four*) and TARGET–distractor Type (*UPRIGHT–inverted*; *UPRIGHT–non-face*; *INVERTED–upright*; *INVERTED–non-face*; *NON-FACE–upright*; *NON-FACE–inverted*). The analysis revealed a significant main effect of Set Size whereby efficiency became poorer as Set Size increased, [$F(4, 156) = 16.65$, $p < .001$, $\eta^2 = 0.30$; Zero, M = 1500 ms, SE = 291 ms; One, M = 1686 ms, SE = 193 ms; Two, M = 1653 ms, SE = 175 ms; Three, M = 2900 ms, SE = 373 ms; Four, M = 2817 ms, SE = 443 ms]. A significant main effect of TARGET–distractor Type was also found [$F(5, 195) = 60.69$, $p < .001$, $\eta^2 = 0.61$], with the poorest efficiency for *UPRIGHT–inverted* (M = 3815 ms, SE = 409 ms) and *INVERTED–upright* (M = 5023 ms, SE = 808 ms) conditions compared to other conditions, (*UPRIGHT–non-face*, M = 973 ms, SE = 139 ms; *INVERTED–non-face*, M = 1140 ms, SE = 160 ms; *NON-FACE–upright*, M = 849 ms, SE = 127 ms; *NON-FACE–inverted*, M = 867 ms, SE = 128 ms). A significant interaction effect between Set Size and TARGET–distractor Type was also found [$F(20, 780) = 8.83$, $p < .001$, $\eta^2 = 0.18$].

Simple main effects revealed significant differences between the Set Sizes only for *UPRIGHT–inverted* and *INVERTED–upright* conditions [$F(4, 936) = 2.61$, $p = .034$, $\eta^2 = 0.01$; $F(4, 936) = 55.66$, $p < .001$, $\eta^2 = 0.19$, respectively]. However, further Tukey's HSD tests found no differences in efficiency between each Set Size for the *UPRIGHT–inverted* condition. Within the *INVERTED–upright* condition, significantly poorer efficiency was found for Set Sizes Three and Four compared to other Set Sizes. No other differences were found between the Set Sizes for the remaining conditions, and detection was consistently efficient.

Simple main effects also revealed significant differences between TARGET–distractor Types at each Set Size. At Set Size Zero [$F(5, 975) = 8.95$, $p < .001$, $\eta^2 =$

0.04], when all items in the display were distractors, *UPRIGHT–inverted* was significantly less efficient than all other conditions. At Set Size One through Three, *UPRIGHT–inverted* and *INVERTED–upright* were detected with the poorest efficiency compared to all other conditions but not each other, [Set Size One, F (5, 975) = 9.89, $p < .001$, $\eta^2 = 0.05$; Set Size Two, F (5, 975) = 7.08, $p < .001$, $\eta^2 = 0.04$; Set Size Three, F (5, 975) = 43.27, $p < .001$, $\eta^2 = 0.18$]. At Set Size Four, when all items are TARGETs, *INVERTED–upright* detection is the least efficient compared to all other conditions, while *UPRIGHT–inverted* is less efficient than *NON-FACE–upright* and *NON-FACE–inverted* only [F (5, 975) = 44.26, $p < .001$, $\eta^2 = 0.18$].

Analyses of IES data within the *UPRIGHT–inverted* condition revealed no significant efficiency costs per additional face in the display. Overall detection efficiency for this condition is poorer compared to other conditions, indicating difficulty distinguishing between these two orientations of faces. Poor detection efficiency is also seen for Set Sizes Three and Four for the *INVERTED–upright* condition where more distractor upright faces are present than target inverted faces. Comparisons between TARGET–distractor Types at each Set Size also reveal a similar pattern. Detection efficiency for conditions that contain Non-Faces – either as targets or distractors – were the most efficient whilst *UPRIGHT–inverted* and *INVERTED–upright* were least efficient. This suggests that discriminating between *Non-Faces* and *Upright* or *Inverted Faces* is easier than discriminating between *Upright* and *Inverted Face* stimuli. This pattern or results further points to an orientation invariant element within the face detection process, as suggested by Experiments 1 and 2. If a difference between *Upright* and *Inverted Face* detection were present, then detecting one stimulus from the other would be an easier task with better efficiency. This was not found in Experiment 3.

What remains evident from Experiment 3 is that when the surrounding environment does not resemble a face (i.e. contains *Non-Faces*), face detection is an efficient and parallel process that does not incur a detection cost per additional face.

This is explored further in the two following analyses, which combine conditions based on the Target as specified to the participant during the experiment or based on the display's visual quality (i.e., when the number and type of items are matched regardless of the condition).

2.4.2.2 Detection Efficiency by Target

Inverse efficiency scores were combined based on common Targets to create an *Upright Faces* condition consisting of *UPRIGHT–inverted* and *UPRIGHT–non-face*; an *Inverted Faces* condition consisting of *INVERTED–upright*; *INVERTED–non-face*; and a *Non-Faces* condition consisting of *NON-FACE–upright*; *NON-FACE–inverted*. Figure 2.5 (A) displays the mean IES score for each new Target Type condition at each Set Size for Experiment 3.

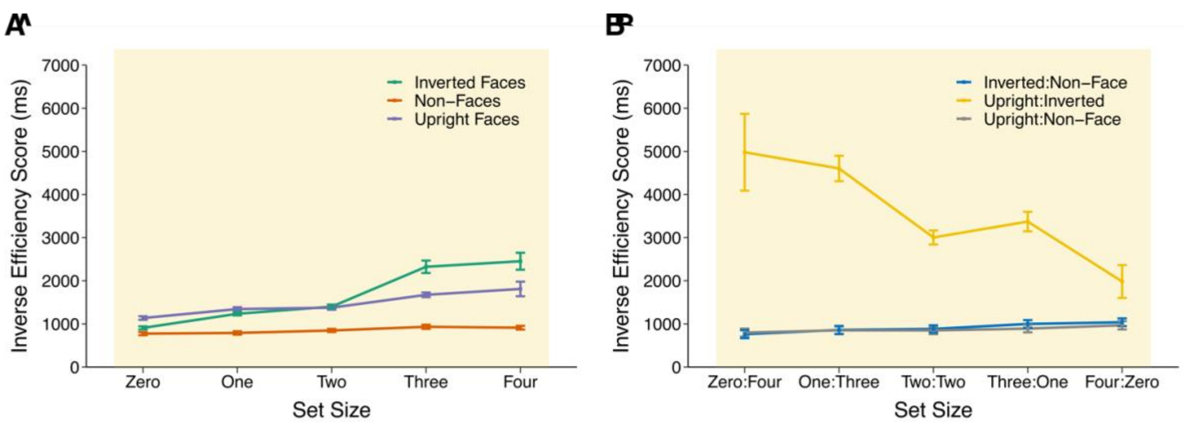

Figure 2.5 Mean IES scores (ms) for Experiment 3 grouped by Target Type (A) and by Display Type (B). Lower scores indicate better efficiency. Error bars show within-subjects standard error (Cousineau, 2005). Yellow shaded areas indicate a significant difference between groups at a particular set size.

To investigate the effect of Target Type IES, data were subjected to a two-way ANOVA with repeated measures of Set Size (*Zero*, *One, Two*, *Three, Four*) and Target Type (*Upright Faces*, *Inverted Faces*, *Non-Faces*). A significant main effect of Set Size was found where efficiency became poorer as Set Size increased [F (4, 156 = 38.65, $p < .001$, $\eta^2 = 0.50$; Zero, M = 939 ms, SE = 38 ms; Two, M = 1208 ms, SE =

44 ms; Three, M = 1644 ms, SE = 82 ms; One, M = 1123 ms, SE = 41 ms; Four, M = 1725 ms, SE = 136 ms]. A significant main effect of Target Type was also found with the greatest efficiency for Non-Faces, [F (2, 78) = 67.13, $p < .001$, $\eta^2 = 0.63$; Upright Faces, M = 1468 ms, SE = 71 ms; Inverted Faces, M = 1664 ms, SE = 93 ms; Non-Faces, M = 851 ms, SE = 41 ms]. The ANOVA also revealed a significant interaction effect between Set Size and Item Type [F (8, 312 = 16.84, $p < .001$, $\eta^2 = 0.30$].

Simple main effects comparing Set Size for each Target Type revealed significant differences for *Upright Face*s [F (4, 468) = 10.52, $p < .001$, $\eta^2 = 0.08$]. Conditions containing *Three* or *Four Upright Faces* were detected with poorer efficiency compared to other Set Sizes, (*Zero Upright Faces*, M = 1136 ms, SE = 42 ms; *One Upright Faces*, M = 1344 ms, SE = 43 ms; *Two Upright Faces*, M = 1376 ms, SE = 47 ms; *Three Upright Faces*, M = 1674 ms, SE = 54 ms; *Four Upright Faces*, M = 1811 ms, SE = 169 ms). Significant differences were also found between Set Sizes within *Inverted Faces*, [F (4, 468) = 67.47, $p < .001$, $\eta^2 = 0.37$]. Conditions containing *Zero Inverted Faces* (M = 907 ms, SE = 37 ms) were detected with the greatest efficiency, followed by *One* (M = 1235 ms, SE = 38 ms) and *Two Inverted Faces* (M = 1400 ms, SE = 48 ms) which were detected with similar efficiency. *Three* (M = 2324 ms, SE = 144 ms) and *Four Inverted Faces* (M = 2453 ms, SE = 197 ms) were detected with the poorest efficiency overall. No Significant differences were found for the different Set Sizes of *Non-Faces*, *Zero Non-Faces*, M = 774 ms, SE = 37 ms; *One Non-Faces*, M = 790 ms, SE = 40 ms; *Two Non-Faces*, M = 848 ms, SE = 36 ms; *Three Non-Faces*, M = 933 ms, SE = 48 ms; *Four Non-Faces*, M = 912 ms, SE = 43 ms).

Simple main effects also revealed significant differences between Target Types at each Set Size, [Zero, F (2, 390) = 4.49, $p = .012$, $\eta^2 = 0.02$; Two, F (2, 390) = 11.57, $p < .001$, $\eta^2 = 0.06$; Three, F (2, 390) = 13.10, $p < .001$, $\eta^2 = 0.06$; One, F (2, 390) = 65.07, $p < .001$, $\eta^2 = 0.25$; Four, F (2, 390) = 80.40, $p < .001$, $\eta^2 = 0.29$]. Across all Set Sizes, *Upright* and *Inverted Faces* targets were detected with significantly poorer efficiency than *Non-Face* Targets. *Upright* and *Inverted Faces* were also not significantly different from each other except at Set Size Three and Four, where they made up the majority of their respective displays.

2.4.2.3 Detection Efficiency by Display

Inverse efficiency scores were combined based on common visual qualities of item types and quantity in the display. For instance, Set Size Three *UPRIGHT–inverted* condition and Set Size One *INVERTED–upright* contain the same item types in the same ratio of 3 Upright Faces: 2 Inverted Faces. Combining conditions based on displays results in three new Item Combination conditions; *Upright:Inverted*, *Upright:Non-Face*, and *Inverted:Non-Face*, as well as new Set Size Ratios of Zero:Four, One:Three, Two:Two, Three:One, Four:Zero. Figure 2.5 (B) displays the mean IES score for each new Item Type condition at each Set Size Ratio for Experiment 3.

To investigate the effect of Display Type IES data were subjected to a two-way ANOVA with repeated measures of Set Size (*Zero:Four, One:Three, Two:Two, Three:One, Four:Zero*) and Item Combination (*Upright:Inverted*, *Upright:Non-Face*, and *Inverted:Non-Face*). A significant main effect of Set Size Ratio was found [F (4, 156 = 4.69, $p < .001$, $\eta^2 = 0.11$]. Detection efficiency improved as Set Size Ratio proceeded in the following pattern: *Zero:Four* (M = 2180 ms, SE = 359 ms), *One:Three*, M = 2105 ms, SE = 159 ms, *Two:Two*, (M = 1579 ms, SE = 111 ms), *Three:One* (M = 1754 ms, SE = 134 ms), *Four:Zero* (M = 1330 ms, SE = 189 ms). A main effect of Item Combination was also found, [F (3, 78 = 105.53, $p < .001$, $\eta^2 = 0.73$]. Detection efficiency was poorest for *Upright:Inverted* (M = 3588 ms, SE = 391 ms), while *Upright:Non-Face* (M = 871 ms, SE = 90 ms) and *Inverted:Non-Face* (M = 909 ms, SE = 90 ms) were detected with similarly greater efficiency. An interaction effect between Set Size and Item Combination was also found [F (8, 312 = 7.12, $p < .001$, $\eta^2 = 0.15$].

Simple main effects comparing Set Size Ratio for each Item Combination revealed significant difference across *Upright:Inverted* condition only [F (4, 468) = 18.67, p = .001, η^2 = 0.14]. Detection efficiencies at *Zero:Four Upright:Inverted* (M = 4980 ms, SE = 890 ms) and *One:Three Upright:Inverted* (M = 4604 ms, SE = 295 ms) were the poorest and not significantly different from each other. At these ratios, Inverted Face stimuli outnumber Upright Face stimuli. Detection efficiencies for *Two:Two Upright:Inverted* (M = 3004 ms, SE = 163 ms) and *Three:One Upright:Inverted* (M = 3372 ms, SE = 226 ms), where Upright Face stimuli are equal to or outnumber Inverted Face stimuli, were intermediate and non-significantly different from each other. The greatest detection efficiency within *Upright:Inverted* was found for *Four:Zero* (M = 1983 ms, SE = 381 ms) which contained four upright faces within the display.

No Set Size differences were found within *Upright:Non-Face* [F (4, 468) = 0.05, p = .995, η^2 = 0.00; *Zero:Four*, M = 798 ms, SE = 92 ms; *One:Three*, M = 851 ms, SE = 90 ms; *Two:Two*, M = 849 ms, SE = 86 ms; *Three:One*, M = 891 ms, SE = 87 ms; *Four:Zero*, M = 968 ms, SE = 96 ms], or within *Inverted:Non-Face* [F(4, 468) = 0.16, p = .960, η^2 = 0.00; *Zero:Four*, M = 761 ms, SE = 93 ms; *One:Three*, M = 861 ms, SE = 93 ms; *Two:Two*, M = 883 ms, SE = 85 ms; *Three:One*, M = 999 ms, SE = 89 ms; *Four:Zero*, M = 1038 ms, SE = 90 ms].

Simple main effects also found significant differences between each Item Combination for the different Set Size Ratios. Across all Set Size Ratios, a clear separation in IES can be seen. *Upright:Inverted* was detected with significantly poorer efficiency, while *Upright:Non-Face* and *Inverted:Non-Face* were consistently detected with similarly greater efficiency [*Zero:Four*, F (2, 390) = 68.56, p = .001, η^2 = 0.26; *One:Three*, F (2, 390) = 54.56, p = .001, η^2 = 0.22; *Two:Two*, F (2, 390) = 17.76, p = .001, η^2 = 0.08; *Three:One*, F (2, 390) = 22.92, p = .001, η^2 = 0.11; *Four:Zero*, F (2, 390) = 3.75, p = .024, η^2 = 0.02].

Across the analyses by Target or Display type, a clear pattern emerges showing that detecting upright face stimuli from inverted face stimuli is a difficult task. This further suggests an orientation invariant element to the detection process. If the face detection mechanism was specialised to upright faces, then an advantage would have been seen when upright faces were the targets, or when they outnumbered other items in the display. But the results of Experiment 3 imply that this is not the case. However, detecting faces, regardless of orientation, from non-face stimuli was a substantially easier task that did not incur detection costs as the number of faces increases. Experiment 3 shows that yet again face detection outperforms non-face detection. Moreover, within 'categorical detection' contexts, where faces and non-faces are present together, faces are still detected in a parallel and efficient manner.

2.5 **General Discussion**

Three main findings emerge from these experiments. First, up to four faces can be detected rapidly and accurately. Second, multiple face detection outperforms multiple non-face detection for up to four items. Third, detection of multiple upright and inverted faces does not differ.

The first two findings concern the quantity, speed, and accuracy of multiple face detection. Experiments 1 and 2 revealed that up to four faces can be detected more quickly and accurately than non-face objects. More than four faces can still be detected, but beyond four, accuracy decreases (Experiment 1), and reaction times increase (Experiment 2) substantially for each additional face. The boundary in detection performance between the smaller and larger Set Sizes is in line with the observed subitizing limit of around four items (Piazza et al., 2011; Trick & Pylyshyn, 1994). This boundary could result from a shared general item detection mechanism; however, the greater performance for faces over non-faces points to some specificity *or* speciality for face stimuli. Efficiency scores from Experiment 3 further support this interpretation. No costs in efficiency were observed when faces are embedded amongst non-faces, or vice-versa, regardless of set size, suggesting that the two stimulus types were easily distinguishable even at the detection stage. Moreover, the non-faces used in Experiment 3 were highly similar phase-scrambled face equivalents that retained the lower-level visual properties of their source faces. Therefore even when non-faces are similar in this sense, multiple target detection was still more efficient for intact faces.

A detection advantage for faces over non-faces has been reported in the limited prior detection literature (Crouzet et al., 2010; Keys et al., 2021; Purcell & Stewart, 1986, 1988; Wardle et al., 2020). However, the current experiments demonstrate for the first time that more than one face can be detected efficiently, and that the face advantage can be sustained over set sizes greater than one. Furthermore, by combining accuracy and reaction time into inverse efficiency scores, Experiment 3 shows that no detection costs are incurred per additional face in the display. Compared to the strict capacity limit of one in the later face processes (Bindemann et al., 2007; Bindemann, Burton, et al., 2005), this pattern of results suggests that the bottleneck in face perception may only constrain processing bandwidth after the detection stage. The results also allude to parallel detection mechanism for up to four faces. However, this study was only designed to establish if a difference in multiple face detection existed compared to multiple non-face detection. Determining the exact face processing bottleneck or assessing the serial-*vs*-parallel nature of face detection (cf. Lewis & Edmonds, 2005; Lewis & Ellis, 2003; Nothdurft, 1993) requires experiments designed to test those ideas specifically.

The third finding refers to an inversion-invariance at the detection stage. Across all three experiments, upright and inverted face stimuli were detected in the same manner regardless of Set Size. This pattern is especially evident in the poor detection efficiency for upright and inverted faces within the same displays in Experiment 3. Inversion-invariance in the detection stage contrasts with poorer performance for inverted faces seen in later face processes (Farah et al., 1995; Yovel & Kanwisher, 2005a). However, it does support the conclusion of the detection template seems insensitive to vertical orientation. Consequently, detection may be driven by a qualitative aspect that is shared by upright and inverted faces but not their non-face comparisons (e.g. retained internal features).

The combination of a face detection advantage that is inversion-invariant suggests a broadly tuned face detection template. It has been suggested that lower-level visual properties may underlie this detection template (Crouzet et al., 2010; Crouzet & Thorpe, 2011), and hence inversion-invariance. But the difference in detection found for upright and inverted faces compared to scrambled faces in Experiment 3 may suggest otherwise. All three stimulus types share the same lower-level visual qualities, such as spatial frequency and colour. Other theories have proposed a role for horizontal "bar-code" energies – the alternating dark and light regions of a face (Dakin & Watt, 2009a; Goffaux & Dakin, 2010) in detection. As well as a role for horizontal eye pairs, which have been found to be sufficient to elicit face detection (Kauffmann et al., 2021; Simpson, Maylott, Mitsven, et al., 2019). Future experiments could assess the role of bar-code representations and horizontal eye pairs by testing the detection of faces that are rotated by 90º rather than 180º. The resulting 'sideways' faces would lack both of the proposed characteristics.

The current series of experiments explored multiple face detection to establish a foundation for future research. Across all three experiments, we found that up to four faces can be detected efficiently compared with non-faces. Unlike later stages of face perception, detection performance was unaffected by face inversion. Our findings raise further questions regarding the capacity limit and location of the face processing bottleneck, the serial-*vs*-parallel nature of face detection, and the qualities of the face detection template that might affect performance.

Chapter 3 – Capacity Limits in Face Detection

This chapter has been published in the following journal, it has been adapted to fit the formatting of this thesis :

Qarooni, R., Prunty, J., Bindemann, M., & Jenkins, R. (2022). Capacity limits in face detection. *Cognition*, *228*, 105227.

3.1 Abstract

Face detection is a prerequisite for further face processing, such as extracting identity or semantic information. Those later processes appear to be subject to strict capacity limits, but the location of the bottleneck is unclear. In particular, it is not known whether the bottleneck occurs before or after face detection. Here we present a novel test of capacity limits in face detection. Across four behavioural experiments, we assessed detection of multiple faces via observers' ability to differentiate between two types of display. *Fixed* displays comprised items of the same type (all faces or all non-faces). *Mixed* displays combined faces and non-faces. Critically, a 'fixed' response requires all items to be processed. We found that additional faces could be detected with no cost to efficiency, and that this capacity-free performance was contingent on visual context. The observed pattern was not specific to faces, but detection was more efficient for faces overall. Our findings suggest that strict capacity limits in face perception occur after the detection step.

3.2 Introduction

Studies of face perception often emphasise the wealth of social information that we derive from faces. However, access to this information is gated by the prior step of face detection, in which the visual system registers the presence of a face. Despite its gatekeeper role, face detection has received little research attention compared to later stages of face perception (e.g. identification; social inferences). Most of the research on face detection concerns algorithm development in computer vision (see Hjelmås & Low, 2001; Kumar et al., 2018 for reviews). As such, the cognitive process of face detection is not well understood.

We follow previous researchers in assuming that face detection involves matching a region of the visual field to a stored face template (Lewis & Ellis, 2003; Robertson et al., 2017; Tsao & Livingstone, 2008). The few psychological studies that have addressed this process have tended to focus on qualitative aspects of the putative template, such as sensitivity to the colour, outline, or spatial layout of the face (Amso et al., 2014; Bindemann & Burton, 2009; Crouzet et al., 2010; Pongakkasira & Bindemann, 2015; Purcell & Stewart, 1986, 1988; Simpson, Maylott, Leonard, et al., 2019; Stein et al., 2012). Less still is known about quantitative aspects of face detection, such as whether multiple faces can be detected at once. In some ways this is a puzzling omission, as quantitative aspects of later face perception processes (e.g. gaze perception, identification, semantic association) have been studied in some detail (Bindemann et al., 2007; Bindemann, Burton, et al., 2005; Bindemann & Burton, 2009; Jenkins et al., 2003).

Several of those studies have recruited the notion of *capacity limits*—the basic observation that not all the available sensory information can be processed at once (Bruckmaier et al., 2020; Lavie & De Fockert, 2003; Norman & Bobrow, 1975). The further claim is that face processing may be subject to its own, face-specific capacity limits (Jenkins et al., 2003). Surprisingly, this limit may be as low as a single face, such that face processing proceeds one face at a time.

The evidence leading to this claim comes from a range of behavioural experiments. For example, patterns of response competition effects (Bindemann, Burton, et al., 2005; Jenkins et al., 2003; Thoma & Lavie, 2013) and repetition priming effects (Bindemann et al., 2007) indicate that processing one face selectively blocks processing of another face. This finding applies not only to later, cognitively deep processes involving extraction of personal identity or semantic information (Bindemann et al., 2007; Bindemann et al., 2005; Jenkins et al., 2003; Thoma & Lavie, 2013), but also to earlier, cognitively shallow processes such as classifying sex (male/female, Bindemann et al., 2005) or gaze direction (left/right, Bindemann & Burton, 2009).

Together, these findings point to a bottleneck early in face processing (i.e. upstream of sex or gaze perception) that constrains face processing downstream of the bottleneck. One possibility is that face detection itself is the bottleneck. This possibility implies strict capacity limits at the detection stage, such that only one face at a time can be acquired from the visual environment (though still allowing rapid serial acquisition). Alternatively, detection itself could be capacity free. This possibility implies that multiple faces can be acquired in parallel, they just cannot be processed in parallel. On this view, the bottleneck occurs when extracting information from faces.

Can we know that more than one face is present? ERP experiments offer some evidence on this point. The amplitude of the N170, a face-selective ERP marker, has been found to increase when multiple faces are presented (Puce et al., 2013). However, as the authors acknowledge, their task of reporting the number of stimuli (1–3) did not require participants to know whether or not the stimuli were faces. Registering the presence of any stimuli could produce the same results.

Several behavioural studies have addressed the related question of whether a target face 'pops out' from surrounding distractors in a visual search task (Brown et al., 1997; Kuehn & Jolicoeur, 1994; Lewis & Edmonds, 2005; Nothdurft, 1993; Treisman & Gelade, 1980). However, these studies have led to conflicting results. For instance, Nothdurft (1993) found that search times increased with the number of distractors (set size), suggesting serial processing. In contrast, Lewis & Edmonds (2005) found equivalently low search times regardless of set size, suggesting parallel processing. We return to this discrepancy in the General Discussion section.

Although visual search can be informative, there are several reasons why it may be unsuitable for probing capacity limits in face detection. First, the task imposes a distinction between target and distractor stimuli. This distinction gives special status to the target category, potentially affecting attentional set (Bindemann et al., 2007; Wolfe & Horowitz, 2004). Second, visual search entails active scanning for the target (Jenkins et al., 2003), whereas everyday face detection often occurs incidentally during passive viewing. Third, and most importantly for the current study, visual search does not lend itself to testing detection of *multiple* faces. As the participant's task is to indicate the presence or absence of a target, search can be terminated when a single target is found, even if other targets are present.

Previous researchers have distinguished two components of face detection that are sometimes conflated — localising and categorising (Bindemann & Lewis, 2013a). The localising component involves searching for a target under spatial uncertainty. Categorisation involves establishing whether or not a stimulus is a face. As we are primarily interested in template matching, our focus here was the categorisation aspect of detection. To probe capacity limits, we sought to assess the cognitive cost of increasing the number of items, while eliminating any cognitive costs associated with localising those items.

To this end, we devised a new task in which all items have equal status. Participants saw face and non-face items in ‘fixed’ displays (all one stimulus type) or ‘mixed’ displays (a combination of both types). To reduce the need for visual scanning, these items appeared at predefined locations surrounding central fixation at low eccentricities (that is, with spatial certainty for addressable coordinates; Garner et al., 2021). For each display, the participants’ task was to indicate whether the items were fixed or mixed. Critically, this task involves assimilating multiple faces. In particular, correct ‘fixed’ responses require each item to match (or not match) a face detection template before a response is made. We take reportability via this fixed/mixed judgement as our detection criterion. Manipulating the type and number of items in the display allows us to estimate per-item detection costs separately for each stimulus type. We take positive cost per item to indicate that detection is capacity limited. Conversely, we take zero cost per item to indicate that detection is capacity free.

3.3 Experiment 4: Two-*vs*-Three Faces and Non-Faces (Dissimilar)

We began by comparing detection efficiency for faces and non-faces presented in ‘fixed’ or ‘mixed’ displays of set size two or three. Non-faces in this experiment were scrambled (phase-shifted) faces that matched the low-level visual energies of the intact face stimuli (Jenkins et al., 2003). For set size two, *Fixed* conditions contained either two faces (FF) or two non-faces (NN), while *Mixed* conditions contained one stimulus of each type (FN or NF, differentiated by spatial layout). Set size three conditions were constructed by adding an extra face or non-face to the display. Thus, *Fixed* conditions contained either three faces (FFF) or three non-faces (NNN), while *Mixed* conditions combined both types of stimuli (FNN and NFF). Participants were asked to decide as quickly and accurately as possible whether each display was *Fixed* (all the same type of stimulus), or *Mixed* (a combination of both stimulus types). Comparing set size two and three allowed us to estimate the effect of an extra display item on these determinations. We expected that capacity-limited face detection should enforce serial processing, resulting in greater efficiency for set size two than for set size three. On the other hand, capacity-free face detection should allow parallel processing, resulting in equivalent efficiency for set size two and for set size three.

3.3.1 Methods

3.3.1.1 Participants

Seventy-seven participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment. Seventeen participants were excluded due to failed attention checks (as described below; exclusion criteria: 2 or more failed checks within the same block, or 3 across the whole experiment) or slow responses (>2.5 SD from the group mean). The final sample (N = 60) comprised 22 females and 38 males (age range 18–73; M=27.95, SD=11.09).

The sample size of 60 participants was selected based on previous face perception studies (Bindemann et al., 2005, 2007; Fysh, 2018; Thoma & Lavie, 2013), with a margin added to account for variations associated with online testing (Anwyl-Irvine et al., 2020). This allowed for greater confidence the experimental design's ability to detect statistically and psychologically meaningful differences between groups. The sample size of 60 participants was consistent throughout the experiments presented in this thesis, with the exception of Experiments 1, 2, and 3 (laboratory testing only) and Experiment 14 (single trial design).

3.3.1.2 Design and Stimuli

Stimuli were generated using a local face bank of 288 faces. The local bank consisted of AI generated faces (Karras et al. & Nvidia, 2018) supplemented with real faces from the MR2 face bank (Strohminger et al., 2016) and other online sources. This local face bank contained an equal distribution of age (younger adults and older adults), sex (male and female), and ethnicity (Asian, Black, and Caucasian; see Prunty et al., (2022) for details of demographic categorisation). Each image was cropped to a 380-pixel wide × 570-pixel high rectangle to create the face items. To create the non-face items, each face was submitted to Fourier phase transformation that randomly scrambled the phase of component spatial frequencies while maintaining overall brightness, contrast, and orientation (Honey et al., 2008). Figure 3.1 shows examples of this manipulation.

To construct Set Size *Two* displays, we used 58 randomly selected faces from the local face bank together with their scrambled face counterparts. For Set Size *Three*, we used 87 faces and scrambled faces. No item was repeated within a display or within a condition. The selected items were randomly allocated to three predetermined locations that formed an upright equilateral triangle around central fixation (nearest contours ~95 pixels apart).

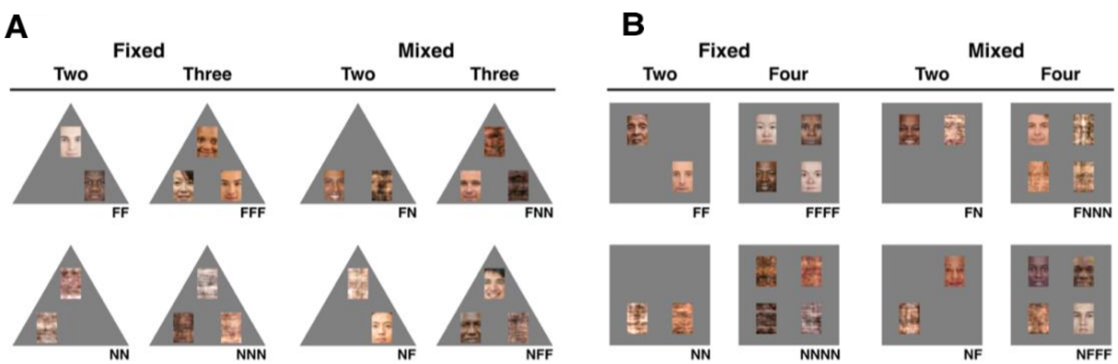

Figure 3.1 Example displays for each condition of Experiment 4 (A) and Experiment 5 (B). Fixed displays contained one type of stimulus (all faces or all non-faces). Mixed displays contained both types. Numbers refer to set sizes. F denotes face, N denotes non-face (scrambled faces in Experiments 4 & 5). Triangle and square segments are for visualisation only. In the actual experiments, the grey background filled the whole screen.

Catch trials for each condition were created for use as attention checks. These catch trials were constructed in a similar manner and used the same spatial layout. Intact faces were substituted with a white vertical rectangle containing a smaller black rectangle in the centre. Scrambled faces were replaced with a phase-scrambled equivalent of these rectangles.

The within-subjects factors of Display Type (*Fixed*, *Mixed*) and Set Size (*Two*, *Three*) were manipulated in a fully counterbalanced 2 × 2 factorial design, resulting in the experimental conditions summarised in Figure 3.1 (A).

The experiment was created and hosted online at Gorilla Experiment Builder (gorilla.sc; Anwyl-Irvine et al., 2020). Participants could access the experiment on any desktop or laptop computer, precluding exact control over screen size. Mobile devices and tablets were excluded.

3.3.1.3 Procedure

Participants were asked to indicate as quickly and accurately as possible whether items in a display were *Fixed* and of the same type (i.e., all face or all non-face), or *Mixed* and a combination of both types (i.e., faces and non-faces together). Each trial started with a fixation cross for 250 ms followed by stimulus displays presented until response. The experiment began with a practice block of 16 trials consisting of two different trials per condition in random order. This was followed by 3 experimental blocks, each consisting of 72 experimental trials (9 trials per conditions) plus 4 catch trials in a random order. Participants were given the opportunity to take short breaks between the blocks. The entire experiment took approximately 10 minutes to complete.

3.3.2 Results and discussion

Overall accuracy for the Fixed/Mixed judgements was 95%, confirming that participants could distinguish between face and non-face stimuli. Trials with reaction times below 150 ms or above 3000 ms were excluded from analysis (0.57% of all trials).

The aims of this chapter require a sensitive metric of efficiency that could account for speed-accuracy trade-offs and measure the efficiency cost per additional item. Consequently, accuracy and reaction time data in Experiments 4 – 7 will be combined to create Linear Integrated Speed Accuracy Scores (LISAS; Vandierendonck, 2017, 2018, 2021). The LISAS metric is specific to within-subjects designs as scores for individual participants are calculated independently from other participants, making it especially appropriate for Chapter 3 experiments. Nonetheless IES metrics remain appropriate for the remaining experiments within this thesis due to their methodological and experimental design which spread speed-accuracy trade-offs across conditions minimizing the impact.

For concision, we combined accuracy and reaction time data to form Linear Integrated Speed-Accuracy Scores (LISAS) which summarise performance in a single efficiency metric while accounting for speed-accuracy trade-offs in responses (Vandierendonck, 2017, 2018, 2021). Separate analyses of accuracy and reaction time measures for each experiment are reported in Supplementary Materials and support the same conclusions.

Figure 3.2 (A) summarises Linear Integrated Speed-Accuracy Scores for each condition in Experiment 1. LISAS data were submitted to a two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Three*) and Display Type (*Fixed*, *Mixed*). This analysis revealed a significant main effect of Set Size, with more efficient detection for *Two* items (M = 768 ms, SE = 9 ms) than for *Three* items (M = 785 ms, SE = 10 ms) overall [$F(1, 59) = 9.16$, $p = .004$, $\eta^2 = 0.31$], and a significant main effect of Display Type, with more efficient detection for *Fixed* displays (M = 756 ms, SE = 10 ms) than for *Mixed* displays (M = 797 ms, SE = 9 ms) overall [$F(3, 177) = 13.08$, $p < .001$, $\eta^2 = 0.18$]. There was also a significant interaction between Set Size and Display Type [$F(3, 177) = 5.33$, $p = .002$, $\eta^2 = 0.08$], reflecting stronger effects of Set Size for *Mixed* displays than for *Fixed* displays.

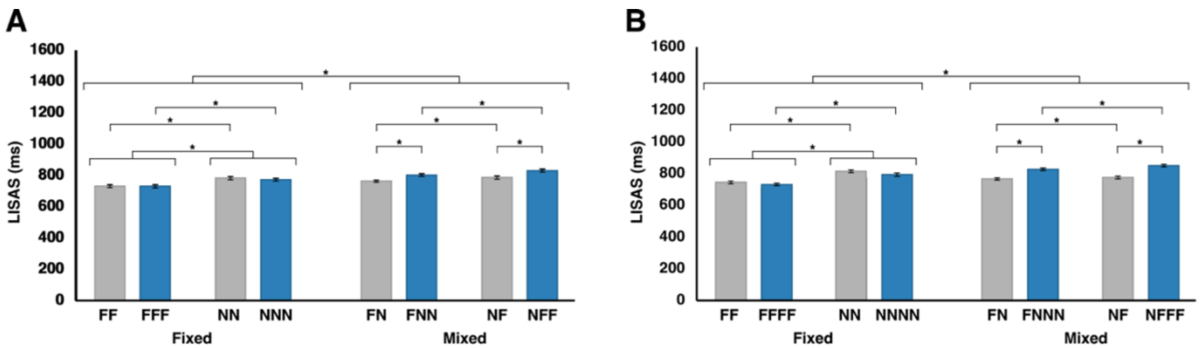

Figure 3.2 Mean linear integrated speed-accuracy scores (LISAS) for each condition in Experiment 4 (A) and Experiment 5 (B) Lower scores indicate better efficiency. F denotes face, N denotes non-face (scrambled faces in Experiments 4 & 5). Error bars show within-subjects standard error (Cousineau, 2005). Brackets indicate significant differences between groups.

For *Fixed* displays, there was no cost incurred by adding a face [FF, M = 734 ms, SE = 10 ms; FFF, M = 732 ms, SE = 11 ms; F (1, 236) = 0.02, *p* = .889, η^2 = 0.00] or a non-face [NN, M = 785 ms, SE = 11 ms; NNN, M = 775 ms, SE = 10 ms; F (1, 236) = 0.76, *p* = .383, η^2 = 0.00]. However, for *Mixed* displays, there was a significant cost to adding either a face [NF, M = 788 ms, SE = 10 ms; NFF, M = 831, SE = 10 ms; F (1, 236) = 13.65, *p* <.001, η^2 = 0.05] or a non-face [FN, M = 766 ms, SE = 7 ms; FNN, M = 803 ms, SE = 8 ms; F (1, 236) = 10.55, *p* = .001, η^2 = 0.04]. The effect of Display Type was significant for Set Size *Two* [F (3, 354) = 5.49, *p* = .001, η^2 = 0.04] and also for Set Size *Three* [F (3, 354) = 15.90, *p* < .001, η^2 = 0.12].

Data from the *Fixed* conditions were submitted to separate t-tests to compare detection of faces and non-faces at each Set Size. Detection was significantly more efficient for faces than for non-faces at Set Size *Two* [*t* (59) = -2.97, *p* = .004, *d* = .38] and at Set Size *Three* [*t* (59) = -2.43, *p*=.018, *d* =.31].

Analysing responses to *Fixed* displays allowed us to compare detection efficiency for *Two* versus *Three* items of the same category (all faces or all non-faces). This comparison revealed no evidence of capacity limits, in the sense that there was no effect of set size: adding an extra item incurred no efficiency cost. However, at both set sizes, detection was more efficient for faces than for non-faces. The next experiment introduces a stronger manipulation of set size.

3.4 Experiment 5: Two-*vs*-Four Faces and Non-Faces (Dissimilar)

To amplify possible effects of set size in this task, we next doubled the magnitude of the set size manipulation. Instead of adding one extra item to the displays (as in Experiment 4), we now added two extra items to the displays. We reasoned that doubling the strength of the set size manipulation should double the size of any latent performance costs.

3.4.1 Methods

3.4.1.1 Participants

Eighty new participants were recruited through Prolific and completed the experiment in exchange for a small payment. Twenty participants were excluded due to failed attention checks (2 or more within the same block, or 3 across the whole experiment) or slow responses (>2.5 SD from the group mean). The final sample (N = 60) comprised 21 females and 39 males (age range 18-49; M=26.10, SD=7.45).

3.4.1.2 Design and Stimuli

The stimuli and catch trials were the same as in Experiment 1 but were now presented in displays of either two items or four items. A total of 68 faces and 68 non-faces were required for each of the Set Size *Two* conditions, and a total of 136 faces and 136 non-faces were required for each of the Set Size *Four* conditions. In each display, the selected items were randomly allocated to four predetermined locations that formed a square around central fixation (nearest contours ~95 pixels; see Figure 3.1b).

3.4.1.3 Procedure

The procedure was the same as in Experiment 4, except that we increased the number of experimental trials. Participants now completed 4 experimental blocks, each consisting of 64 experimental trials (8 trials per conditions plus 3 catch trials) in a random order. The entire experiment took approximately 10 minutes to complete.

3.4.2 Results and Discussion

Overall accuracy for the Fixed/Mixed judgements was 95%, confirming that participants could distinguish between face and non-face stimuli. Trials with reaction times below 150 ms or above 3000 ms were excluded from analysis (1.24% of all trials).

Figure 3.2 (B) summarises Linear Integrated Speed-Accuracy Scores for each condition in Experiment 2. LISAS data were submitted to a two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Four*) and Display Type (*Fixed*, *Mixed*). This analysis revealed a significant main effect of Set Size, with more efficient detection for *Two* items (M = 762 ms, SE = 8 ms) than for *Four* items (M = 786 ms, SE = 9 ms) overall [$F(1, 59) = 15.36$, $p < .001$, $\eta^2 = 0.21$], and a significant main effect of display type, with more efficient detection for *Fixed* displays (M = 756 ms, SE = 8 ms) than for *Mixed* displays (M = 793 ms, SE = 8 ms) overall [$F(3, 177) = 25.44$, $p < .001$, $\eta^2 = 0.30$]. There was also a significant interaction between Set Size and Display Type [$F(3, 177) = 18.97$, $p < .001$, $\eta^2 = 0.26$], reflecting stronger effects of Set Size for *Mixed* displays than for *Fixed* displays.

For *Fixed* displays, there was no cost incurred by adding two extra faces [FF, M = 728 ms, SE = 8 ms; FFFF, M = 715 ms, SE =7 ms; $F(1, 236) = 1.37$, $p = .243$, $\eta^2 = 0.01$], or two extra non-faces [NN, M = 800 ms, SE = 8 ms; NNNN, M = 779 ms, SE = 10 ms; $F(1, 236) = 3.34$, $p = .069$, $\eta^2 = 0.01$]. However, for *Mixed* displays, there was a significant cost to adding either two faces [NF, M = 764 ms, SE = 9 ms;

NFFF, M = 838, SE = 8 ms; F (1, 236) = 41.06, $p < .001$, $\eta^2 = 0.15$] or two non-faces [FN, M = 756 ms, SE = 6 ms; FNNN, M = 814 ms, SE = 7 ms; F (1, 236) = 25.86 $p < .001$, $\eta^2 = 0.15$]. The effect of Display Type was significant for Set Size *Two* [F (3, 354) = 10.74, $p < .001$, $\eta^2 = 0.08$] and also for Set Size *Four* [F (3, 354) = 35.14, $p < .001$, $\eta^2 = 0.23$].

Data from the *Fixed* conditions were submitted to separate t-tests to compare detection of faces and non-faces at each set size. Detection was significantly more efficient for faces than for non-faces at Set Size *Two*, [t (59) = -5.12, $p < .001$, $d = .66$] and at Set Size *Four* [t (59) = -4.71, $p < .001$, $d = .61$].

As in Experiment 4, comparing detection efficiency for items of the same category revealed no evidence of capacity limits, despite the fact that the magnitude of the set size manipulation was now doubled. Apparently, acquiring four items was no less efficient than acquiring two. We again found that detection was more efficient for faces than for non-faces. Given that stimulus–template match depends on properties of the stimulus, we next asked what differences between faces and non-faces are required for efficient detection of multiple faces.

3.5 Experiment 6: Two-*vs*-Three Faces and Non-Faces (Similar)

The high detection efficiency for faces in Experiments 4 and 5 cannot have been due to their low-level visual energies, given the different result for the phase-shifted faces. However, phase-shifted faces do not control for contour and structural information in the intact face. In the next experiment, we used inverted faces as the comparison stimuli instead. For consistency across experiments, we refer to these inverted face stimuli as 'non-faces'. Given that inverted faces are identical to upright faces in every respect except orientation, we expected the visual distinction between face and non-face stimuli to be less clear, potentially reducing task efficiency.

3.5.1 Methods

3.5.1.1 Participants

Seventy-three participants, who were recruited online via Prolific, completed the experiment in exchange for a small payment. Thirteen participants were excluded due to failed attention checks (2 or more within the same block, or 3 across the whole experiment) or slow responses (>2.5 SD from the group mean). The final sample (N = 60) comprised 26 females and 34 males (age range 19–61; M = 30.22; SD = 9.11).

3.5.1.2 Design and Stimuli

The design was the same as for Experiment 4 except that the intact and scrambled faces were replaced with upright and inverted faces. The upright faces were taken from the original bank of 288 faces and segmented from the background using the InterFace software package (Kramer et al., 2017). The resulting images were resized to 570 pixels high × 380 pixels wide. The non-face stimuli were created by rotating the upright faces 180º in the picture plane. As in Experiment 1, stimuli were combined to create *Fixed* and *Mixed* displays of Set Size *Two* and Set Size *Three*. Example displays are shown in Figure 3.3 (A).

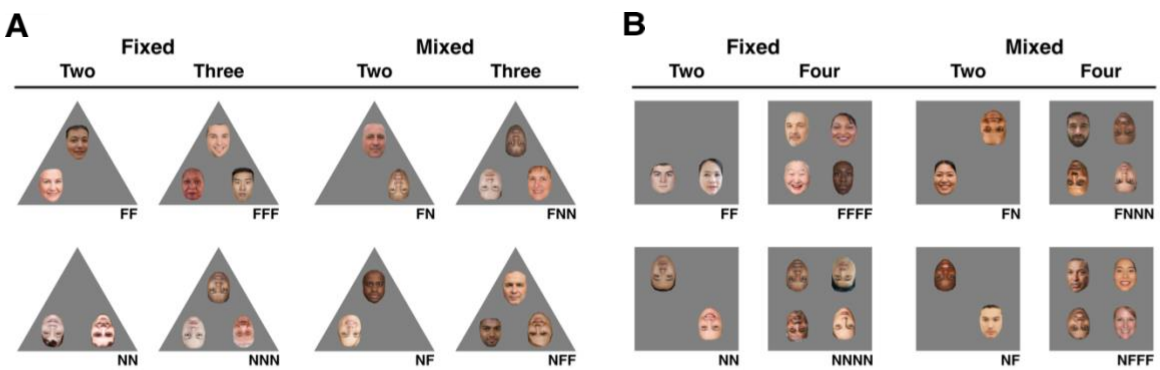

Figure 3.3 Example displays for each condition of Experiment 6 (A) and Experiment 7 (B). Fixed displays contained one type of stimulus (all faces or all non-faces). Mixed displays contained both types. Numbers refer to set sizes. F denotes face, N denotes non-face (scrambled faces in Inverted Faces 6 & 7). Triangle and square segments are for visualisation only. In the actual experiments, the grey background filled the whole screen.

Catch trials for each condition were again created for use as attention checks. These catch trials were constructed in a similar manner and used the same spatial layout as the experimental trials. White circles containing upward- or downward-pointing black arrows were used in place of upright or inverted faces, respectively.

3.5.1.3 Procedure

The procedure was the same as in Experiment 4 and took approximately 10 minutes to complete.

3.5.2 Results and Discussion

Overall accuracy for the Fixed/Mixed judgements was 94%, again confirming that participants could distinguish between face and non-face stimuli, despite their increased similarity in this experiment. Trials with reaction times below 150 ms or above 3000 ms were excluded from analysis (1.6% of all trials).

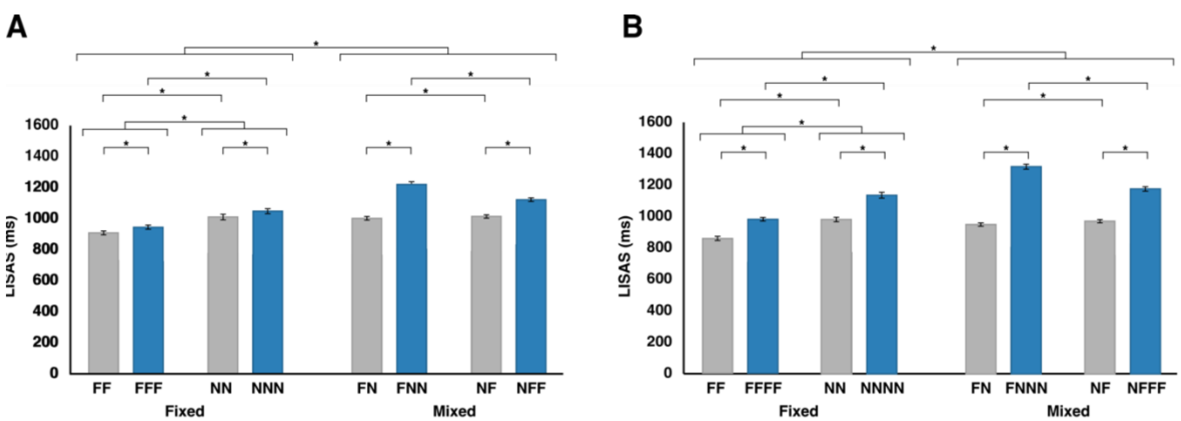

Figure 3.4 Mean linear integrated speed-accuracy scores (LISAS) for each condition in Experiment 6 (A) and Experiment 7 (B). Lower scores indicate better efficiency. F denotes face, N denotes non-face (inverted faces in Experiments 6 & 7). Error bars show within-subjects standard error (Cousineau, 2005). Brackets indicate significance between groups.

Figure 3.4 (A) summarises Linear Integrated Speed-Accuracy Scores for each condition in Experiment 3. LISAS data were submitted to a two-way ANOVA with the repeated measures factors of Set Size (*Two, Three*) and Display Type (*Fixed*, *Mixed*) This analysis revealed a significant main effect of Set Size, with more efficient detection for *Two* items (M = 991 ms, SE = 12 ms) than for *Three* items (M = 1092 ms, SE = 14 ms) overall [F (1, 59) = 127.29, $p < .001$, $\eta^2 = 0.68$], and a significant main effect of display type, with more efficient detection for *Fixed* displays (M = 985 ms, SE = 15 ms) than for *Mixed* displays (M = 1098 ms, SE = 13 ms) overall [F (3, 177) = 31.68, $p < .001$, $\eta^2 = 0.39$]. There was also a significant interaction between Set Size and Display Type [F (3, 177) = 27.65, $p < .001$, $\eta^2 = 0.32$], reflecting stronger effects of Set Size for *Mixed* displays than for *Fixed* displays.

For *Fixed* displays, there was a significant cost incurred by adding an extra face [FF M = 915 ms, SE = 12 ms; FFF, M = 952 ms, SE =14 ms; F (1, 236) = 4.89, $p = .028$, $\eta^2 = 0.03$] or an extra non-face [NN, M = 1018 ms, SE = 18 ms; NNN, M = 1056 ms, SE = 17 ms; F (1, 236) = 4.99, $p = .028$, $\eta^2 = 0.02$].

For *Mixed* displays, there was also a significant cost to adding an extra face [NF, M = 1022 ms, SE = 11 ms; NFF, M = 1131 ms, SE = 12 ms; F (1, 236) = 41,60, $p < .001$, $\eta^2 = 0.15$] or an extra non-face [FN, M = 1009 ms, SE = 11 ms; FNN, M = 1230 ms, SE = 16 ms; – F (1, 236) = 171.91, $p < .001$, $\eta^2 = 0.42$]. The effect of Display Type was significant for Set Size *Two* [F (3, 354) = 11.04, $p < .001$, $\eta^2 = 0.09$] and also for Set Size *Three* [F (3, 354) = 58.77, $p < .001$, $\eta^2 = 0.33$].

Data from the *Fixed* conditions were submitted to separate t-tests to compare detection of faces and non-faces at each set size. Detection was significantly more efficient for faces than for non-faces at Set Size *Two*, [t (59) = -4.33, $p < .001$, $d = .56$], and at Set Size *Three* [t (59) = -4.25, $p < .001$, $d = .55$].

Analysis of the *Fixed* conditions again allowed us to compare detection efficiency for *Two* versus *Three* items of the same category (all faces or all non-faces). Unlike Experiment 4, this comparison revealed clear evidence of capacity limits, in the sense that there was a significant effect of set size: adding an extra item incurred a substantial efficiency cost, as expected from the reduced distinction between face and non-face stimuli. In keeping with the preceding experiments, detection was again more efficient for faces than for non-faces. This face advantage was seen not only in the *Fixed* conditions, but also in the *Mixed* conditions, where adding a face to a display incurred a smaller cost than adding a non-face.

3.6 Experiment 7: Two-*vs*-Four Faces and Non-Faces (Similar)

To better understand the effects of set size seen in Experiment 6, we doubled the magnitude of the set size manipulation (similar to Experiment 5). If performance is capacity-limited for these new stimuli, such that each extra item incurs its own performance cost, then doubling the number of extra items (from 1 to 2) should double the cost.

3.6.1 Methods

3.6.1.1 Participants

Sixty-eight new participants were recruited through Prolific and completed the experiment in exchange for a small payment. Eight participants were excluded due to failed attention checks (2 or more within the same block, or 3 across the whole experiment) or slow responses (>2.5 SD from the group mean). The final sample (N = 60) comprised 23 females and 37 males (age range 18–56; M = 27.25; SD = 8.93).

3.6.1.2 Design and Stimuli

The design, stimuli, and catch trials, were the same as in Experiment 6, except that the display items comprised either two or four items that were randomly allocated to four predetermined locations that formed a square around central fixation (see Figure 3.3 B).

3.6.1.3 Procedure

The procedure was the same as in Experiment 5 and took approximately 10 minutes to complete.

3.6.2 Results and Discussion

Overall accuracy for the Fixed/Mixed judgements was 92%, again confirming that participants could distinguish between face and non-face stimuli. Trials with reaction times below 150 ms or above 3000 ms were excluded from analysis (0.97% of all trials).

Figure 3.4 (B) summarises Linear Integrated Speed-Accuracy Scores for each condition in Experiment 4. LISAS data were submitted to a two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Four*) and Display Type (*Fixed*, *Mixed*). This analysis revealed a significant main effect of Set Size, with more efficient detection for *Two* items (M = 981 ms, SE = 14 ms) than for *Four* items (M = 1202 ms, SE = 15 ms) overall [$F(1, 59) = 313.66$, $p < .001$, $\eta^2 = 0.84$], and a significant main effect of display type, with more efficient detection for *Fixed* displays (M = 1031 ms, SE = 15 ms) than for *Mixed* displays (M = 1132 ms, SE = 13 ms) overall [$F(3, 177) = 57.87$, $p < .001$, $\eta^2 = 0.50$]. There was also a significant interaction between Set Size and Display Type [$F(3, 177) = 46.62$, $p < .001$, $\eta^2 = 0.44$], reflecting stronger effects of Set Size for *Mixed* displays than for *Fixed* displays.

For *Fixed* displays, there was a significant cost incurred by adding two extra faces [FF, M = 895 ms, SE = 14 ms; FFFF, M = 1021 ms, SE =13 ms; $F(1, 236) = 44.00$, $p < .001$, $\eta^2 = 0.16$], or two extra non-faces [NN, M = 1023 ms, SE = 14 ms; NNNN, M = 1184 ms, SE = 119 ms; $F(1, 236) = 70.82$, $p < .001$, $\eta^2 = 0.23$]

Similarly, for *Mixed* displays, there was a significant cost to adding two extra faces [NF, M = 1014 ms, SE = 11 ms; NFFF, M = 1227 ms, SE = 14 ms; $F(1, 236) = 124.93$, $p < .001$, $\eta^2 = 0.35$] or two extra non-faces [FN, M = 991 ms, SE = 11 ms; FNNN, M = 1375 ms, SE = 17 ms; $F(1, 236) = 403.01$, $p < .001$, $\eta^2 = 0.63$]. The effect of Display Type was significant for Set Size *Two* [$F(3, 354) = 15.29$, $p < .001$, $\eta^2 = 0.11$] and for Set Size *Four* [$F(3, 354) = 93.51$, $p < .001$, $\eta^2 = 0.44$].

Data from the *Fixed* conditions were submitted to separate t-tests to compare detection of faces and non-faces at each set size. Detection was significantly more efficient for faces than for non-faces at Set Size *Two*, [$t(59) = -6.34$, $p < .001$, $d = .82$] and at Set Size *Four* [$t(59) = -6.41$, $p < .001$, $d = .83$].

As with Experiment 6, the results of Experiment 7 are consistent with capacity-limited performance in which each additional display item contributes to performance costs. Detection was again more efficient for faces than for non-faces. This face advantage was observed across all experimental conditions.

3.7 General Discussion

Our experiments reveal at least four principles of multiple face detection. First and foremost, we show that viewers can capture additional faces at no extra cost. Second, cost-free capture is contingent on visual context. Third, this facility is not specific to faces. Fourth, it is efficient for faces. We address each of these points in turn.

Can viewers know that more than one face is present? Yes. Figure 3.5 summarises the cost per additional item for fixed/mixed judgements across experiments. Experiment 4 shows that adding an item to the display incurred no efficiency cost. Even when the number of items was doubled from two to four (Experiment 5), we found no impact on efficiency. In fact, in none of these cases was the per-item cost even numerically positive. We conclude that the visual system can acquire multiple faces concurrently, at least over the range of 2–4 faces tested here. This finding is consistent with the ERP observation that multiple faces can enhance the N170 (Puce et al., 2013). However, the current task allows us to draw more specific conclusions. Unlike a simple numerosity task, the fixed/mixed task used here required participants to discern whether or not the seen items were intact upright faces.

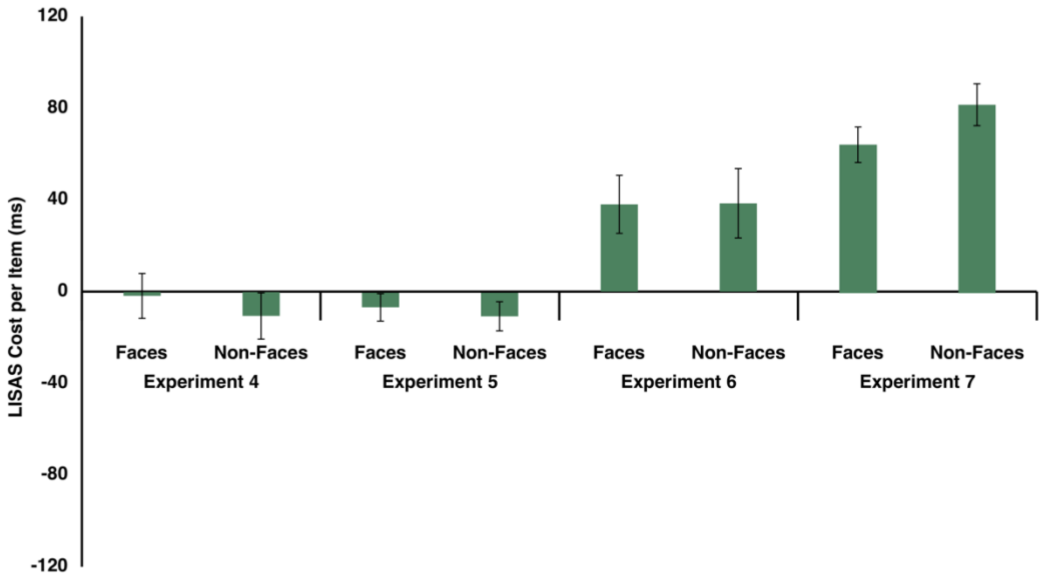

Figure 3.5 Summary of cost per additional item for Fixed conditions (face and non-face) in Experiments 4 – 7.

The observed multi-item capacity at the detection stage of face processing contrasts with surprisingly strict capacity limits seen for later stages of face processing. Response competition experiments requiring judgements of sex, eye direction, and semantic information have repeatedly found that processing one face precludes processing another face (Bindemann et al., 2007; Bindemann, Burton, et al., 2005; Burton & Bindemann, 2009; Jenkins et al., 2003). The cognitively earlier task of face detection apparently evades this strict limit of one. Taken together, these findings help to locate the putative bottleneck in face processing. We suggest that a processing bottleneck occurs *after* the detection step, such that coarse face/non-face discriminations may be conducted in parallel, but before further face information is extracted, such that finer discriminations among faces must be conducted in series. Future experiments could modify the method described here to test the upper bound of multiple face detection—in particular, whether detection capacity exceeds our maximum of four items. We expect that there will be some limit to the number of faces that can be detected concurrently, not least because overall visual bandwidth is limited. Establishing these upper limits for face detection will require careful experimentation, as increasing the number of display items necessitates increasing eccentricity, increasing crowding, or reducing item size, all of which can affect general visual discrimination. One interesting possibility relates to *subitizing*—rapid and accurate enumeration of up to 4 items (Kaufman & Lord, 1949; Piazza et al., 2011; Pylyshyn, 2004). Although classic demonstrations of subitizing relied on simple visual objects (e.g. dots) as stimuli, more recent work has established subitizing-like phenomena for complex objects, including human figures (Railo et al., 2016). The fixed/mixed task is clearly different from subitizing, as it requires participants to know whether or not display items are of the same type, rather than just enumerating them. Even so, the 4-item span associated with subitizing provides theoretical motivation to test larger set sizes in future experiments. For now, equivalent detection efficiency over the range of 2 to 4 items shows that face detection is not subject to a strict capacity limit of just one face.

The comparison between Experiments 4 & 5 (in which the non-faces were scrambled faces) and Experiments 6 & 7 (in which the non-faces were inverted faces) underscores the importance of visual context in determining task performance (see Figure 3.5). Although multiple face detection *can* proceed in parallel (Experiments 4 & 5), whether or not it actually *does* proceed in parallel depends on other factors—in this case, the nature of other items in the display (Experiments 6 & 7). This contingency is useful, as it suggests a way to probe what counts as a face to the visual system. Establishing that scrambled faces are not faces incurred no cognitive cost in this task, implying that an appropriate distribution of visual energies is not itself sufficient to match the face template. In contrast, establishing that inverted faces are not faces did incur a cognitive cost, implying that the spatial organisation of those visual energies makes a meaningful difference. We suggest that an inverted face matches (or partially matches) the face detection template whereas a scrambled face does not. As a result, upright and inverted faces take additional cognitive resources to sort out, apparently one at a time.

This interpretation fits with the broader notion that discriminations among stimuli that activate face processing (e.g. identification; extraction of social signals) are serial in nature. It also suggests a candidate behavioural marker of face template matching. If stimuli that give rise to serial processing in this task engage the template, and stimuli that give rise to parallel processing do not, it should be possible to characterise the 'receptive field' of the face template by varying the non-face stimuli in this task. Further iterations of the present task, with different kinds of carefully-designed distractor items, could reveal what counts (and does not count) as a face to the visual system.

It may seem counterintuitive to emphasise non-face visual content as a determinant of face perception, especially as so much previous work emphasises facial appearance. However, the distinction between face and non-face is key to the early perceptual step of face detection. From this perspective, we should expect performance to be determined as much by the rest of the visual environment as it is by faces themselves. The more closely visual properties of the environment resemble visual properties of faces (cf. scrambled faces in Experiments 4 & 5, inverted faces in Experiments 6 & 7), the more demanding the face/non-face discrimination becomes (Duncan & Humphreys, 1989; Lewis & Edmonds, 2005). This basic insight suggests that it will be difficult to generalise from detection experiments based on isolated faces only to face detection in the real world. It also suggests a principled means to reconcile seemingly discrepant findings in the literature. Visual search studies that have reported 'pop-out' for a target face have avoided presenting other face-like information in their displays (e.g. Lewis & Edmonds 2005, Experiment 7). Those that found no pop-out did present other face-like information (notably inverted faces, e.g. Lewis & Edmonds, 2005, Experiment 2; Nothdurft, 1993; Brown et al., 1997) This distinction among studies of visual search for faces echoes more general findings in visual search. Search is most efficient when targets and distractors are dissimilar and displays contain homogeneous distractors; search becomes less efficient when target–distractor similarity increases irrespective of display heterogeneity (Roper et al., 2013). However, all of those studies relied specifically on a visual search task, in which the experimenters define faces as targets, display items must be localised, and a maximum of one target face is present. The current experiments emphasise the relation between face and non-face material in a very different task, in which no target category is defined, localisation is not required, and multiple faces are acquired from the visual environment simultaneously.

Our final two points concern whether or not faces are 'special' in this situation. We note that response patterns across experiments were qualitatively similar for face and non-face displays. In Experiments 4 and 5, comparison of fixed conditions revealed no evidence of capacity limits for either stimulus category, in that adding extra items incurred no efficiency cost. In Experiments 6 and 7, the same comparison revealed set-size costs for both stimulus categories. As such, we make no claims concerning *qualitative* differences between multiple face detection and multiple stimulus detection generally. However, *quantitative* differences between categories were both clear and consistent. Across all four experiments, responses to faces were more efficient than responses to non-faces. This apparent face advantage accords with previous studies of face detection. For example, detection of upright intact faces has been shown to be more efficient than detection of other objects, pareidolic faces, and even faces with rearranged internal features (Crouzet et al., 2010; Keys et al., 2021; Purcell & Stewart, 1986, 1988; Stein et al., 2012). In our view, there are many possible explanations for this apparent face advantage. The current experiments were not designed to disentangle them. Instead, we conclude that multiple faces can be detected concurrently, implying that the bottleneck in face processing follows the detection step, rather than preceding or coinciding with detection. Whether multiple faces actually are detected concurrently in a particular situation can depend on other aspects of the visual scene.

Chapter 4 – Multiple Face Detection

This chapter has been submitted for publication, it has been adapted to fit the formatting of this thesis :

Qarooni, R., Prunty, J., Bindemann, M., & Jenkins, R. (2022). Multiple Face Detection. *(Submitted)*.

4.1 Abstract

We often encounter several faces at once in complex social scenes. However, previous research on face detection has mainly focused on single targets and blank backgrounds. In the current study, participants reported the number of target items (upright, inverted, or scrambled faces) embedded in scenes of varying complexity. Across manipulations of exposure duration (Experiments 8 & 9), background complexity (Experiments 10 & 11), and target orientation (Experiments 12 & 13), we found rapid parallel detection for up to four faces in complex scenes. Performance was equivalent for upright and inverted faces, but less efficient for sideways faces, possibly reflecting a preference for vertical symmetry. In Experiment 14, we assessed spontaneous multiple face detection in a large-scale single-trial design ($N = 1388$) using 700 photographs of social groups. Presenting groups of up to six people allowed us to test the upper limits of multiple face detection. Beyond four faces, enumeration errors exceeded correct responses and were mainly underestimates. We conclude that humans can detect up to 4 ± 1 faces in complex scenes, efficiently and concurrently. Our findings illuminate an early process in social cognition that is often neglected. They also show that presenting multiple search targets can enrich our understanding of visual performance.

4.2 Introduction

Humans readily form social groups of various sizes (Dunbar, 1998; Dunbar & Spoors, 1995; Zhou et al., 2005). It follows that when we encounter humans, we often encounter several at once rather than isolated individuals. Although work on social processes has examined group dynamics at different time scales (McGrath et al., 2000), very little has looked at the earliest moments of a social encounter, when we first become aware that other people are present.

Can we become aware of several people concurrently, or do we register each person in turn? One of the most reliable visual indicators of company is face detection. Pioneering studies on spatial orienting have found that a face in the visual field tends to capture attention (Bindemann, Burton, et al., 2005; Bindemann & Burton, 2008), and can be distinguished from non-face stimuli in as little as 100 ms (Crouzet et al., 2010; Crouzet & Thorpe, 2011). However, virtually all psychological literature on face detection has examined single faces (Bindemann & Lewis, 2013; Lewis & Edmonds, 2005). Detection of multiple faces has been almost entirely overlooked. While there is established literature on detecting multiple visual items (Katzin et al., 2019; Piazza et al., 2011; Riggs et al., 2006), much of it concerns numerical cognition and has not presented faces as detection targets.

Face detection is thought to occur by matching regions of the visual environment to a stored face template (Lewis & Ellis, 2003; Robertson et al., 2017; Tsao & Livingstone, 2008). Previous research on face detection has focused on the specific qualitative properties involved during the template matching procedure. For efficient detection, a face template seems to require the correct integration of natural colour, shape, and internal spatial information. Manipulating face properties beyond these specific parameters has been shown to impair the detection process (Amso et al., 2014; Bindemann & Burton, 2009; Bindemann & Lewis, 2013; Crouzet &

Thorpe, 2011; Lewis & Edmonds, 2005; Pongakkasira & Bindemann, 2015; Purcell & Stewart, 1986, 1988; Stein et al., 2012). More recently, the role of the eyes has also been proposed as an important cue to face detection. Eye pairs alone have been shown to elicit rapid and accurate detection. However, an intact face representation with all its features was still the most efficient detection template (Kauffmann et al., 2021; Omer et al., 2019). Much of this previous work explored face detection in laboratory settings through visual search or absent/present tasks of a single face in a plain backgrounds (but see (Bindemann & Lewis, 2013; Di Giorgio et al., 2012; Lewis & Edmonds, 2005; Pongakkasira & Bindemann, 2015) for important exceptions). However, face detection in everyday life often involves multiple faces embedded in a complex visual environment. Introducing this quantitative dimension raises an important question: is multiple face detection in naturalistic scenes a serial process or a parallel process?

In a recent behavioural study (Qarooni et al., 2022), we addressed capacity limits in face detection, asking whether viewers can know that more than one face is present. Participants were instructed to decide whether or not a set of display items all belonged to the same category (all intact faces or all scrambled faces) or two different categories (both intact and scrambled faces). Judgements were equally efficient for 2, 3, or 4 faces, demonstrating that several faces can be detected simultaneously (parallel processing). However, changing the scrambled faces to inverted faces introduced measurable per-item detection costs (serial processing), even when the intact faces remained the same. This observation suggests that surrounding visual context can determine the efficiency of multiple face detection.

The displays presented in that study were tightly controlled, in the tradition of psychophysical experiments. While a psychophysical approach can be appropriate when probing cognitive architecture, it does not capture the variability in visual demands that we experience outside laboratory settings. In the current study, we sought to establish the impact of different viewing conditions on multiple face detection. Previous work based on non-face targets suggests that the nature of the visual background is key here. For example, Wolfe et al. (2002) argued that segmenting objects from naturalistic backgrounds occurs in a single-preattentive step, only shifting to item-by-item selection when the background is visually similar to the targets. In a more recent study, Railo et al. (2016) found that increasing visual complexity of the background (from a blank canvas to a natural scene), slowed the parsing of biologically relevant stimuli from scenes. The stimuli used by Railo et al. (2016) depicted human figures in various poses, but importantly for the current study, faces were not always visible. Both studies found a similar pattern. Up to three items could be detected in parallel, and additional items beyond that incurred a detection cost.

Wolfe et al. (2002) and Railo et al. (2016) suggest an essential role of visual background context in general detection. However, there are several reasons why the situation may be different for face detection. First, many studies have concluded that faces are a special category for the perceptual system, and may compete especially strongly for attentional resources (Farah et al., 1998; Tsao & Livingstone, 2008; Valentine, 1988; Yin, 1969a; Yovel & Kanwisher, 2005b). Second, later stages of face processing are known to be strictly capacity limited (one face at a time), implying a bottleneck at some earlier stage (Bindemann et al., 2007). Third, our recent experiments demonstrate parallel detection for up to four faces, but only presented stimuli against blank backgrounds (Qarooni et al., 2022). No studies have examined multiple face detection in naturalistic scenes.

Here we ask how visual properties of search targets and search background affect capacity limits in face detection. In Experiments 8–13, participants were presented with displays containing 1, 2, 3, or 4 search targets embedded in different visual surrounds. Their task was to indicate, as quickly and accurately as possible, how many targets were present. In Experiments 8 & 9, we presented Upright, Inverted, or Scrambled Faces in real scenes, and measured detection efficiency for different exposure durations. In Experiments 10 & 11 we manipulated the visual complexity and meaningfulness of the search backgrounds. In Experiments 12 & 13, we manipulated target type and orientation to probe the tuning of the face detection template. All of the preceding experiments (8–13) involved superimposing isolated search targets onto unrelated scenes. The resulting displays were artificial in the sense that they lacked the coherence of everyday social scenes. In Experiment 14, we tackled these limitations by presenting 700 social group photographs to 1400 participants in a large-scale single-trial study. Replacing the previous search task with a retrospective enumeration judgement also allowed us to rule out possible effects of advance instructions or practice.

4.3 Experiment 8: Upright, Inverted & Scrambled in Real Scenes (250 ms)

The first experiment compares detection efficiency for *Upright, Inverted* and *Scrambled* targets presented at Set Sizes of *One*, *Two*, *Three*, and *Four* in real complex scenes. All items in a given scene were of the same target type, and participants were tasked with reporting how many targets they saw as quickly and accurately as possible. Comparing performance across target types should reveal any categorial differences in detection efficiency. Comparing across set sizes should allow us to estimate the cognitive cost imposed by each target.

4.3.1 Methods

4.3.1.1 Participants

Eighty-two participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment. Twenty-two participants were excluded due to failed attention checks (2 or more within a block, or 3 across the entire experiment) or slow responses (>2.5 SD from the group mean). The final sample (N = 60) comprised 22 females, 36 males, and 2 who preferred not to answer (age range 19 – 58; M = 25.14, SD = 6.97).

4.3.1.2 Design and Stimuli

Stimuli were generated from a local face bank of 288 faces with an equal distribution of age (young and old adults), sex (male and female), and ethnicity (Asian, Black, and Caucasian; see Prunty et al., (2022) for details of demographic categorisation). Face images were collected from AI generated faces (Karras et al. & Nvidia, 2018) supplemented with real faces from the MR2 face bank (Strohminger et al., 2016) and other online sources.

To create the face stimuli, the original face images were cropped to outline using the InterFace software package (Kramer et al., 2017), removing the extraneous background. The resultant images were then resized to 380 pixels wide × 570 pixels high. Inverted face stimuli were created by rotating face stimuli 180° in the picture plane.

Scrambled face stimuli were created by cropping the same original face images to a 380 pixels wide × 570 pixels high rectangle. These cropped images were then submitted to Fourier phase transformation, randomly scrambling the phase of component spatial frequencies while maintaining overall brightness, contrast, and orientation (Honey et al., 2008).

A bank of 300 scene images was used as complex naturalistic backgrounds (see Prunty et al., (2022) for details). All scenes were of real places that contained no faces or people. The bank consisted of 50 images for each of six scene categories: School, Garage, Home, Office, Restaurant, and Shop.

The same procedure was used to generate displays for the *Upright*, *Inverted*, and *Scrambled* conditions. For instance, 180 images were randomly selected from the face stimuli set without repetition to create Upright conditions. Twelve scene images from each of the six scene categories we also randomly selected without repetition resulting in a total of 72 scenes allocated to the *Upright* condition. To avoid extreme peripheral presentations, 25 predetermined locations within the central 75% of the scene were chosen, and 1, 2, 3, or 4 faces were allocated to these locations depending on Set Size. A total of 18 experimental trials were generated for each target and evenly distributed across three blocks. In addition, four practice face trials (one for each target) were generated similarly. The same procedure was repeated for the *Inverted* and *Scrambled* conditions, with no item or scene being repeated across the whole experiment. Example displays can be seen in Figure 4.1(A).

A total of 12 attention check trials were also created, one for each Target in each of the three blocks. These are similar to experimental trials but contain blue circles (average size equal to average target size) on a grey background.

The within-subjects factors of Target Type (*Upright, Inverted, Scrambled*) and Set Size (*One, Two*, *Three, Four*) were manipulated in a fully counterbalanced 3 × 4 factorial design.

The experiment was hosted online at Gorilla Experiment Builder (Gorilla.sc; Anwyl-Irvine et al., 2020). Participants could access the experiment on any desktop or laptop computer resulting in variable screen sizes. Mobile devices and tablets were excluded.

4.3.1.3 Procedure

Participants were shown an example of each target type (*Upright*, *Inverted*, *Scrambled*) and were instructed that each display would contain targets of one type only (e.g. all *Upright* or all *Inverted*). For the participant, *Upright*, *Inverted*, and *Scrambled* targets were labelled Type A, Type B, and Type C, respectively. Participants were also instructed that they may see between 1 and 4 of the target types in the display. For each trial, the participants' task was to indicate, as quickly and accurately as possible, how many of the specified items they saw in the display (numbered keypress response). Each trial began with a blank screen shown for 100 ms, then a fixation cross for 250 ms, and another 100 ms blank screen. The stimulus display was then presented for 250 ms, followed by a text prompt that remained on screen until response. The experiment began with a block of 12 practice trials presented in random order. Three experimental blocks followed this practice block, each comprising 72 experimental trials and 4 attention check trials presented in random order. The experiment lasted an average of 10 minutes in total.

4.3.2 Results and Discussion

Trials with a reaction time below 150 ms or above 3000 ms were excluded (< 0.01% of all trials). For concision, accuracy and reaction times for each trial were combined in a single Inverse Efficiency Score (IES; Bruyer & Brysbaert, 2011; Townsend & Ashby, 1978). Separate analyses of accuracy and reaction time measures are provided in the Supplementary Materials and support the same conclusions.

Inverse Efficiency Scores for each condition are shown in Figure 4.1 (B). The IES data were submitted to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [$F(3, 177) = 19.53$, $p < .001$, $\eta^2 = 0.25$], with greatest detection efficiency for *One* item (M = 1212 ms, SE = 104 ms), intermediate efficiency for *Two* (M = 2017 ms, SE = 186 ms) and *Three* items (M = 1875 ms, SE = 153 ms), and poorest efficiency for *Four* items (M = 2705 ms, SE = 238 ms). There was also a significant main effect of Target Type [$F(2, 118) = 113.66$, $p < .001$, $\eta^2 = 0.66$], with better efficiency for *Upright* targets (M = 1077 ms, SE = 87 ms) and *Inverted* targets (M = 1156 ms, SE = 85 ms) than for *Scrambled* targets (M = 3625 ms, SE = 338 ms). The interaction effect between Set Size and Target Type was also significant [$F(6, 354) = 23.03$, $p < .001$, $\eta^2 = 0.28$].

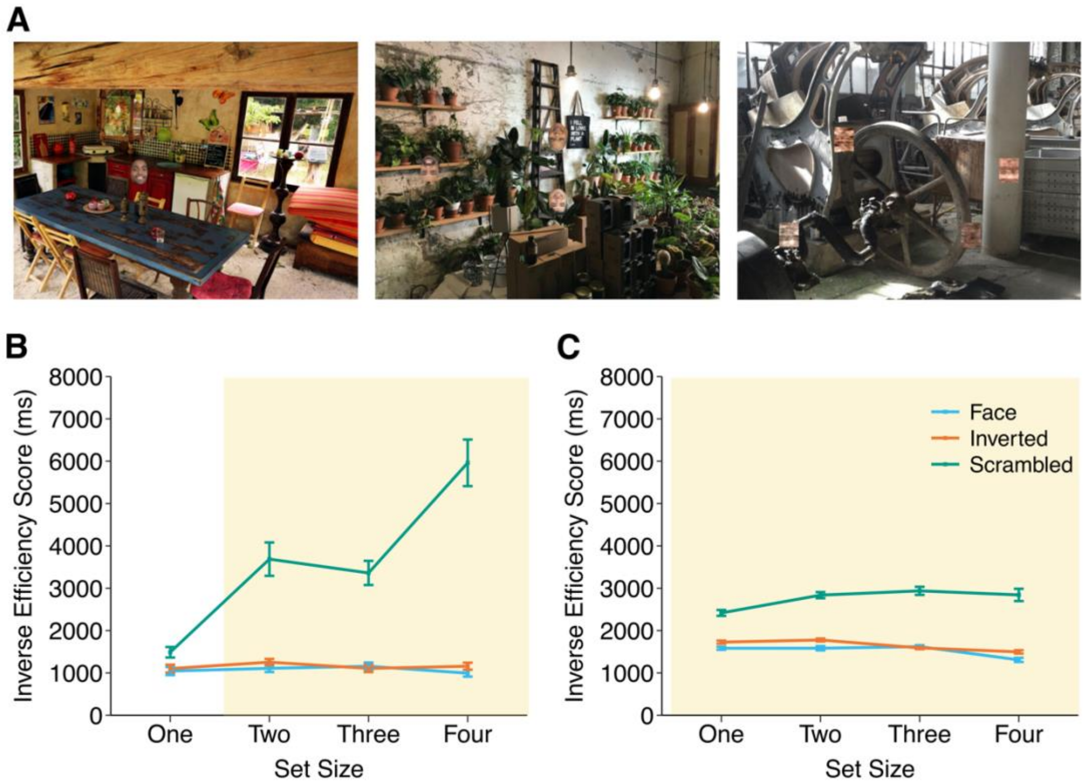

Figure 4.1 Example displays used in Experiments 8 and 9 (A). Mean IES as a function of Set Size, shown separately for Upright, Inverted, and Scrambled targets for Experiment 8 (B) and Experiment 9 (C). Lower scores indicate better efficiency. Error bars show within-subjects standard error (Cousineau, 2005). Yellow shaded area indicates a significant difference Upright and Inverted targets compared to Scrambled targets at a particular set size.

Simple main effects revealed no significant effect of Set Size in the *Upright* condition [One Upright, M = 1045 ms, SE = 95 ms; Two Upright, M = 1109 ms, SE = 87 ms; Three Upright, M = 1159 ms, SE = 87 ms; Four Upright, M = 995 ms, SE = 81 ms; F (3, 531) = 0.10, p = .960, η^2 = 0.00] or the *Inverted* condition [One Inverted, M = 1103 ms, SE = 92 ms; Two Inverted, M = 1254 ms, SE = 77 ms; Three Inverted, M = 1107 ms, SE = 86 ms; Four Inverted, M = 1159 ms, SE = 82 ms; F (3, 531) = 0.09, p = .963, η^2 = 0.00]. In contrast, there was a highly significant effect of Set Size in the *Scrambled* condition [One Scrambled, M = 1489 ms, SE = 124 ms; Two Scrambled, M = 3687 ms, SE = 394 ms; Three Scrambled, M = 3361 ms, SE = 287 ms; Four Scrambled, M = 5961 ms, SE = 550 ms; $F (3, 531) = 65.00$, $p < .001$, $\eta^2 = 0.27$].

There was no significant effect of Target Type at Set Size One [$F (2, 472) = 1.06$, $p = .349$, $\eta^2 = 0.00$]. However, *Upright* and *Inverted* targets were detected more efficiently than *Scrambled* targets at every other Set Size, [Set Size Two, $F (2, 472) = 38.07$, $p < .001$, $\eta^2 = 0.14$; Set Size Three, $F (2, 472) = 30.04$, $p < .001$, $\eta^2 = 0.11$; Set Size Four, $F (2, 472) = 144.39$, $p < .001$, $\eta^2 = 0.38$].

No per-item cost was found over the tested range for *Upright* and *Inverted* targets. This flat set size function suggests parallel detection for upright and inverted faces, even in cluttered scenes. In contrast, increasing the number of *Scrambled* targets strongly affected detection efficiency, indicating serial detection for these items. One possible reason for this disparity is that the short exposure duration (250 ms) made it difficult for participants to segment *Scrambled* targets from the surrounding scene. In the next experiment, we test this possibility by removing the exposure limit.

4.4 Experiment 9: Upright, Inverted & Scrambled in Real Scenes (unlimited exposure)

Experiment 9 replicates Experiment 8 and presents displays until participants' response to investigate the effect of display presentation times on face detection efficiency.

4.4.1 Methods

4.4.1.1 Participants

Sixty-three participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment.

Three participants were excluded due to failed attention checks (2 or more within a block, or 3 across the entire experiment) or slow responses (> 2.5 SD from the group mean). The final sample (N = 60) comprised 28 females and 32 males (age range 19 – 56; M = 28.63, SD = 8.20).

4.4.1.2 Design and Stimuli

The design and stimuli were identical to Experiment 8 (See Figure 4.1 (A) for example displays).

4.4.1.3 Procedure

The procedure in Experiment 9 is identical to Experiment 8, except displays are now presented until response. The response prompt was presented once before each block rather than after each trial.

4.4.2 Results and Discussion

No trials were excluded as none fell below the 150 ms reaction time exclusion criteria. Accuracy and reaction times were again used to calculate IES data. Separate analyses of accuracy and reaction time measures are provided in the Supplementary Materials and support the same conclusions as IES.

Figure 4.1 (C) shows IES data for each condition for Experiment 9. IES data were submitted to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [$F(3, 177) = 4.83$, $p = .003$, $\eta^2 = 0.08$], with good detection efficiency for *One* item (M = 1909 ms, SE = 47 ms), intermediate efficiency for *Two* (M = 2066 ms, SE = 52 ms) and *Three* items (M = 2050 ms, SE = 53 ms), and greater efficiency was for *Four* items (M = 1882 ms, SE = 78 ms). The analysis also revealed a significant main effect of Target Type was also found [$F(2, 118) = 239.50$, $p < .001$, $\eta^2 = 0.80$]. *Upright* Targets (M = 1524 ms, SE =

42 ms) were detected with the greatest efficiency, followed by intermediate efficiency for *Inverted* Targets (M = 1648 ms, SE = 35 ms), and poorest efficiency for *Scrambled* Targets (M = 2759 ms, SE = 95 ms). The interaction effect between Set Size and Target Type was also significant [$F (6, 354) = 9.69$, $p < .001$, $\eta^2 = 0.14$].

Simple main effects revealed a significant effect of Set Size across the *Upright* condition. However, as in Experiment 8, no detection costs were incurred for each additional face in the display, [One Upright, M = 1593 ms, SE = 35 ms; Two Upright, M = 1584 ms, SE = 50 ms; Three Upright, M = 1622 ms, SE = 33 ms; Four Upright, M = 1308 ms, SE = 48 ms, $F (3, 531) = 5.17$, $p = .002$, $\eta^2 = 0.03$]. The same pattern was seen in the *Inverted* condition, a significant effect of Set Size was found but no efficiency costs were incurred per additional inverted face in the display [One Inverted, M = 1727 ms, SE = 36 ms; Two Inverted, M = 1778 ms, SE = 35 ms; Three Inverted, M = 1590 ms, SE = 29 ms; Four Inverted, M = 1497 ms, SE = 39 ms, $F (3, 531) = 4.02$, $p < .001$, $\eta^2 = 0.02$]. As with Experiment 8, simple main effects revealed significant detection costs for additional items in the *Scrambled* condition [$F (3, 531) = 13.25$, $p < .001$, $\eta^2 = 0.07$]. One Scrambled Target (M = 2417 ms, SE = 69 ms) was detected more efficiency compared to Two (M = 2837 ms, SE = 72 ms), Three (M = 2939 ms, SE = 96 ms), and Four (M = 2842 ms, SE = 145 ms) Scrambled Target.

No significant effect of Target Type was found at any Set Size between *Upright* and *Inverted* Targets. However, both Target Types were significantly more efficient than *Scrambled* Targets at all Set Sizes [Set Size One, $F (2, 472) = 44.87$, $p < .001$, $\eta^2 = 0.16$; Set Size Two, $F (2, 472) = 102.68$, $p < .001$, $\eta^2 = 0.30$; Set Size Three, $F (2, 472) = 133.62$, $p < .002$, $\eta^2 = 0.36$; Set Size Four, $F (2, 472) = 157.87$, $p < .001$, $\eta^2 = 0.40$].

The results from Experiment 9 replicate those of Experiment 8. While overall IES data was higher than in Experiment 8, no detection costs were found for *Upright* or *Inverted* Targets at any Set Size. Both conditions were detected efficiently for up to four items tested here. But again, detection costs and poor efficiency were found for *Scrambled* Targets.

The unlimited display presentation times in Experiment 9 suggest that the pattern of results seen in the previous experiment is not due to brief or limited exposure duration. Instead, it indicates that our visual system can efficiently discern the presence of multiple faces from real and complex scenes regardless of time limits.

4.5 Experiment 10: Upright, Inverted & Scrambled in Blank Scenes (250 ms)

In Experiment 10, we explore the effect of removing background complexity and meaningfulness on face detection efficiency. We repeat Experiment 8, but now *One, Two, Three* or *Four Upright, Inverted,* and *Scrambled* Targets are embedded in blank grey backgrounds. Removing all background information should reduce the need to segment *Upright* and *Inverted* Targets from the surrounding visual context facilitating detection. For the same reason, *Scrambled* Targets detection is also expected to be efficient. If detection for all Target Types is facilitated, then this would suggest an essential role of the background for the face detection process.

4.5.1 Methods

4.5.1.1 Participants

Sixty-four participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment. Four participants were excluded due to failed attention checks (2 or more within a block, or 3 across the entire experiment) or slow responses (> 2.5 SD from the group mean). The final sample (N = 60) comprised 31 females and 29 males (age range 19 – 62; M = 30.19, SD = 10.22).

4.5.1.2 Design and Stimuli

The design and to-be-detected Targets were identical to previous experiments. The background scenes were now replaced with blank grey backgrounds to remove both meaningfulness and complexity. Example Displays are presented in Figure 4.2 (A).

4.5.1.3 Procedure

The procedure in Experiment 10 is identical to Experiment 8 with displays presented for 250 ms.

4.5.2 Results and Discussion

Trials with a reaction time below 150 ms and above 3000 ms were excluded from the final analysis (< 0.01% of all trials). Accuracy and reaction times were again used to calculate IES data. Separate analyses of both constituent measures are provided in the Supplementary Materials and support the same conclusions as IES.

Figure 4.2 (B) shows IES data for each condition for Experiment 10. The IES data were submitted to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [$F(3, 177) = 51.09$, $p < .001$, $\eta^2 = 0.46$]. Detection efficiency was greatest for *One* item (M = 485 ms, SE = 7 ms), and became poor as Set Size increased (*Two,* M = 509 ms, SE = 8 ms; *Three* items M = 593 ms, SE = 9 ms; *Four*, M = 597 ms, SE = 10 ms). No significant main effect of Target Type was found [*Upright Faces,* M = 547 ms, SE = 42 ms; *Inverted Faces,* M = 545 ms, SE = 35 ms; *Scrambled Faces* M = 546 ms, SE = 95 ms; $F(2, 118) = 0.10$, $p = .905$, $\eta^2 = 0.00$]. However a significant interaction effect between Set Size and Target Type was revealed [$F(6, 354) = 3.67$, $p = .002$, $\eta^2 = 0.06$].

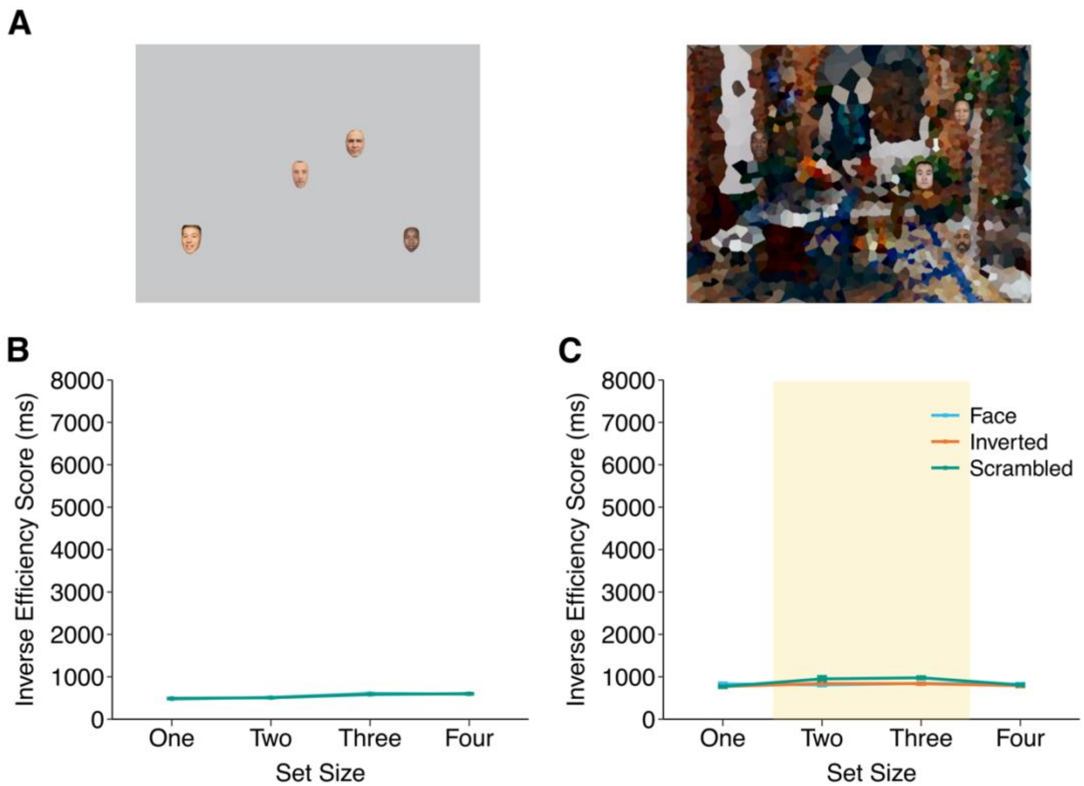

Figure 4.2 Example displays used in Experiments 10 (A – Left) and 11 (A – Right). Mean IES as a function of Set Size, shown separately for Upright, Inverted, and Scrambled targets for Experiment 10 (B) and Experiment 11 (C). Lower scores indicate better efficiency. Error bars show within-subjects standard error (Cousineau, 2005). Yellow shaded area indicates a significant difference Upright and Inverted targets compared to Scrambled targets at a particular set size.

Simple main effects across Set Size revealed a different pattern of results to prior experiments. A significant effect of Set Size was found in the *Upright* condition, as Set Size increased detection efficiency for *Upright* Targets became poorer, and three and four Targets were detected with similar poor efficiency [One Upright, M = 474 ms, SE = 6 ms; Two Upright, M = 515 ms, SE = 8 ms; Three Upright, M = 611 ms, SE = 10 ms; Four Upright, M = 588 ms, SE = 9 ms, [$F (3, 531) = 45.17$, $p < .001$, $\eta^2 = 0.20$]. A significant difference in detection efficiency across Set Size was also for *Inverted* Targets. One Inverted (M = 588 ms, SE = 8 ms) and Two Inverted Targets (M = 506 ms, SE = 7 ms) were detected with similarly greater efficiency than Three Inverted (M = 587 ms, SE = 8 ms) and Four Inverted Targets (M = 598 ms, SE

= 11 ms) which were detected with similarly poorer efficiency [F (3, 531) = 34.32, p < .001, $\eta^2 = 0.16$]. This same significant effect across Set Size was found for Scrambled Targets [One Scrambled, M = 493 ms, SE = 8 ms; Two Scrambled, M = 505 ms, SE = 8 ms; Three Scrambled, M = 583 ms, SE = 9 ms; Four Scrambled, M = 597 ms, SE = 11 ms; F (3, 531) = 34.09, p < .001, $\eta^2 = 0.16$].

Simple main effects revealed no differences between the *Upright*, *Inverted*, and *Scrambled* conditions at each Set Size, except Set Size Three. Three Inverted and Three Scrambled Targets were detected more efficiently than Three Upright Targets [Set Size One, F (2, 472) = 2.54, p = .080, $\eta^2 = 0.01$; Set Size Two, F (2, 472) = 0.69, p = .503, $\eta^2 = 0.00$; Set Size Three, F (2, 472) = 5.90, p = .003, $\eta^2 = 0.02$; Set Size Four, F (2, 472) = 1.60, p = .203, $\eta^2 = 0.01$].

Overall detection efficiency for Experiment 10 was greater than previous experiments for all Item Types; however, a different pattern of results was observed. When background complexity and meaningfulness are removed, upright face detection seems to incur a small (within ~90 ms) but significant detection cost per additional item. Both ‘non-face’ stimuli (inverted and scrambled faces) incurred a detection cost only from Two items to Three items. At the same time, no overall differences were found between *Upright, Inverted* and *Scrambled* conditions except at Set Size 3. It appears all Target Types were easier to segment from a background when meaningfulness and complexity were removed. These results support the importance of including real and/or complex backgrounds in face detection studies. The difference between face and ‘non-face’ information is vital within face detection. Therefore it would be difficult to deduce a pattern or capacity limits from experiments of isolated faces in plain backgrounds.

4.6 Experiment 11: Upright, Inverted, Scrambled in Voronoi Scenes (250 ms)

Experiment 11 builds on the previous experiment and investigates face detection within Voronoi Scenes. These Voronoi Scenes remove background meaningfulness but retain reduced background complexity. Testing face detection under these conditions enables us to understand whether the meaningfulness of the scene is essential or if complexity, regardless of meaning, is needed for segmenting faces from the background.

4.6.1 Methods

4.6.1.1 Participants

Sixty-nine participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment. Nine participants were excluded due to failed attention checks (2 or more within a block, or 3 across the entire experiment) or slow responses (> 2.5 SD from the group mean). The final sample (N = 60) comprised 38 females, 21 males, and 1 who preferred not to answer (age range 19 – 50; M = 25.65, SD = 7.36).

4.6.1.2 Design and Stimuli

The design and to-be-detected targets were identical to previous experiments, except Voronoi scenes equivalents were now used as background context. Voronoi scenes are created by drawing a boundary around a coplanar point connecting all points closer to it than points further away (Roos, 1993). This creates a tessellation of irregular polygon cells that reduces the overall meaningfulness and complexity of a scene but retains the same colour information. An example of a Voronoi scene is shown in Figure 4.2 (A) alongside other example displays from Experiment 11. Compared to a categorisation task of intact scenes (Mean accuracy 75.36%, SE = 6.95%), Voronoi scenes were classified into their appropriate categories with significantly lower accuracy (M = 28.40%, SE = 2.28%), [t (29) = 2.05, $p < .001$] suggesting lower overall complexity and meaningfulness.

4.6.1.3 Procedure

The procedure in Experiment 11 is identical to Experiment 8 with displays presented for 250 ms.

4.6.2 Results and Discussion

Trials with a reaction time below 150 ms and above 3000 ms were excluded from the final analysis (< 0.01% of all trials). IES data was calculated using accuracy and reaction time. Separate analyses of both measures are provided in the Supplementary Materials and support the same conclusions as IES.

Figure 4.2 (C) shows IES data for each condition for Experiment 11. IES data were submitted to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 2.85, $p = .039$, $\eta^2 = 0.05$]. Detection efficiency was greatest for *One* item (M = 485 ms, SE = 30 ms), and became poorer at *Two* items (M = 868 ms, SE = 38 ms), and then *Three* items (M = 886 ms, SE = 25 ms), followed by slightly higher efficiency for *Four* items (M = 807 ms, SE = 29 ms). A significant main effect of Target Type was also found with greater detection efficiency for *Inverted* (M = 810 ms, SE = 28 ms) and *Upright Faces* (M = 827 ms, SE = 26 ms), and poorest efficiency for *Scrambled Faces* (M = 877 ms, SE = 37 ms), [F (2, 118) = 4.57, $p = .012$, $\eta^2 = 0.07$]. A significant interaction effect between Set Size and Target Type was also revealed [F (6, 354) = 4.54, $p < .001$, $\eta^2 = 0.07$].

Simple main effects revealed no significant differences in detection efficiency across Set Size for *Upright* Targets [One Upright, M = 831 ms, SE = 30 ms; Two

Upright, M = 810 ms, SE = 29 ms; Three Upright, M = 843 ms, SE = 21 ms; Four Upright, M = 824 ms, SE = 25 ms; $F(3, 531) = 0.16$, $p = .920$, $\eta^2 = 0.00$], or *Inverted* Targets, [One Inverted, M = 771, SE = 33; Two Inverted, M = 842 ms, SE = 29 ms; Three Inverted, M = 837 ms, SE = 23 ms; Four Inverted, M = 790 ms, SE = 32 ms; $F(3, 531) = 1.09$, $p = .352$, $\eta^2 = 0.01$]. However, *Scrambled* Targets showed differences in detection efficiency across Set Size, with Two (M = 951 ms, SE = 61 ms) and Three (M = 977 ms, SE = 30 ms) Scrambled Targets detected with a similarly poorer efficiency than One (M = 771 ms, SE = 26 ms) and Four (M = 808 ms, SE = 31 ms) Scrambled Targets [$F(3, 531) = 9.09$, $p < .001$, $\eta^2 = 0.05$].

Simple main effects between Target Type at each Set Size revealed no differences at Set Size One [$F(2, 472) = 1.69$, $p = .186$, $\eta^2 = 0.01$]. At both Set Size Two and Set Size Three no differences were found for *Upright* and *Inverted* conditions, however both conditions were significantly more efficient than the *Scrambled* condition [Set Size Two, $F(2, 472) = 7.50$, $p = .001$, $\eta^2 = 0.03$; Set Size Three, $F(2, 472) = 8.62$, $p < .001$, $\eta^2 = 0.04$]. No differences between item Types were found at Set Size Four [$F(2, 472) = 0.40$, $p = .669$, $\eta^2 = 0.00$].

Experiment 11 showed a similar pattern of results to Experiments 8 and 9 for upright and inverted faces. When background complexity is reduced and meaningfulness is removed, no detection costs are incurred for either upright or inverted faces. Nor are any differences found across Set Sizes. However, scrambled faces detection did incur a cost as Set Size increased. These results suggest that visual complexity is important for face detection, further reinforcing the need for incorporating real and/or complex scenes in face detection studies.

4.7 Experiment 12: Scrambled and Sideways in Real Scenes (250 ms)

Across the previous four Experiments, detection efficiency for upright and inverted faces is highly similar. This pattern of results is in contrast to what is observed in later processes of recognition and identification, which are severely impaired by face inversion (Farah et al., 1998; Yin, 1969b). The inversion effect within face detection is contested, with some work specifying that an upright face template is needed for efficient detection (Purcell & Stewart, 1986, 1988). While other work, including our own, argues for an orientation invariant element to the detection template (Bindemann & Burton, 2008; Devue et al., 2012; Qarooni et al., 2022). More evidence also suggests that detection can still occur with just a pair of eyes present (Kauffmann et al., 2021; Omer et al., 2019).

Experiment 12 addresses how manipulating orientation affects the face detection template during multiple face detection. Specifically, we disturb the horizontal eye-pair orientation by comparing *Sideways* Targets *(90° left* or *90° right oriented faces)* with *Scrambled* Targets. If face detection is specialised for detecting upright-oriented faces, then detection efficiency for *Sideways Faces* is expected to be similar to *Scrambled Faces*. If face detection is orientation invariant and it is merely the presence of a pair of eyes in any orientation, then *Sideways Face* detection should not incur a detection cost as Set Size increases and should be more efficient than *Scrambled Faces*

4.7.1 Methods

4.7.1.1 Participants

Sixty-eight participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment. Eight participants were excluded due to failed attention checks (2 or more within a block, or 3 across the entire experiment) or slow responses (> 2.5 SD from the group mean). The final sample (N = 60) comprised 40 females and 20 males (age range 19 – 43; M = 25.73, SD = 6.09).

4.7.1.2 Design and Stimuli

The design and scene stimuli were identical to Experiments 8. Stimuli for *Sideways* conditions consisted of two separate conditions of faces rotated 90° left and 90° right in the picture plane. The same *Scrambled* Targets condition as previous Experiments was used here. Example displays for Experiment 12 can be seen in Figure 4.3 (A).

4.7.1.3 Procedure

The procedure in Experiment 12 is identical to Experiment 8 with displays presented for 250 ms.

4.7.2 Results and Discussion

Trials with a reaction time below 150 ms and above 3000 ms were excluded from the final analysis (< 0.01% of all trials). IES data was calculated using accuracy and reaction time. Separate analyses of both measures are provided in the Supplementary Materials and support the same conclusions as IES.

The separate 90° left and 90° right oriented face conditions were collapsed into the single *Sideways Faces* condition.

Figure 4.3 (B) shows IES data for each condition for Experiment 12. IES data were submitted to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Sideways, Scrambled*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 21.44, $p < .001$, $\eta^2 = 0.27$]. Detection efficiency was greatest for *One* item (M = 1313 ms, SE = 202 ms), followed by *Two* items (M = 2137 ms, SE = 229 ms), and then *Three* items (M = 3421 ms, SE = 474 ms), and *Four* items (M = 4003 ms, SE = 330 ms). A significant main effect of Target Type was found with greater detection efficiency for *Sideways* Targets (M = 1276 ms, SE = 186 ms) compared to *Scrambled* Targets (M = 4161 ms, SE = 431 ms), [F (1, 59) = 61.94, $p < .001$, $\eta^2 = 0.51$]. A significant interaction effect between Set Size and Target Type was also found [F (3, 177) = 21.48, $p < .001$, $\eta^2 = 0.27$].

Simple main effects revealed no significant differences in detection efficiency across Set Size for *Sideways* Targets [One Sideways, M = 1198 ms, SE = 199 ms; Two Sideways, M = 1337 ms, SE = 191 ms; Three Sideways, M = 1251 ms, SE = 179 ms; Four Sideways, M = 1320 ms, SE = 175 ms, [F (3, 354) = 0.03, $p = .993$, $\eta^2 = 0.00$]. Differences in detection efficiency for *Scrambled* Targets were found, with efficiency becoming poorer as set size increases, but Three and Four Scrambled Targets were detected similarly (One Scrambled, M = 771 ms, SE = 26 ms; Two Scrambled, M = 771 ms, SE = 26 ms; Three Scrambled, M = 771 ms, SE = 26 ms; Four Scrambled, M = 771 ms, SE = 26 ms; F (3, 354) = 42.89, $p < .001$, $\eta^2 = 0.27$).

Simple main effects between Target Type at each Set Size revealed no differences at Set Size One [F (1, 236) = 0.16, $p = .688$, $\eta^2 = 0.00$]. However efficiency for the *Sideways* condition was significantly better than *Scrambled* condition at each following Set Size [Set Size Two, F (1, 236) = 7.73, $p = .006$, $\eta^2 = 0.03$; Set Size Three, F (1, 236) = 56.76, $p < .001$, $\eta^2 = 0.19$; Set Size Four [F (1, 236) = 86.81, $p < .001$, $\eta^2 = 0.27$].

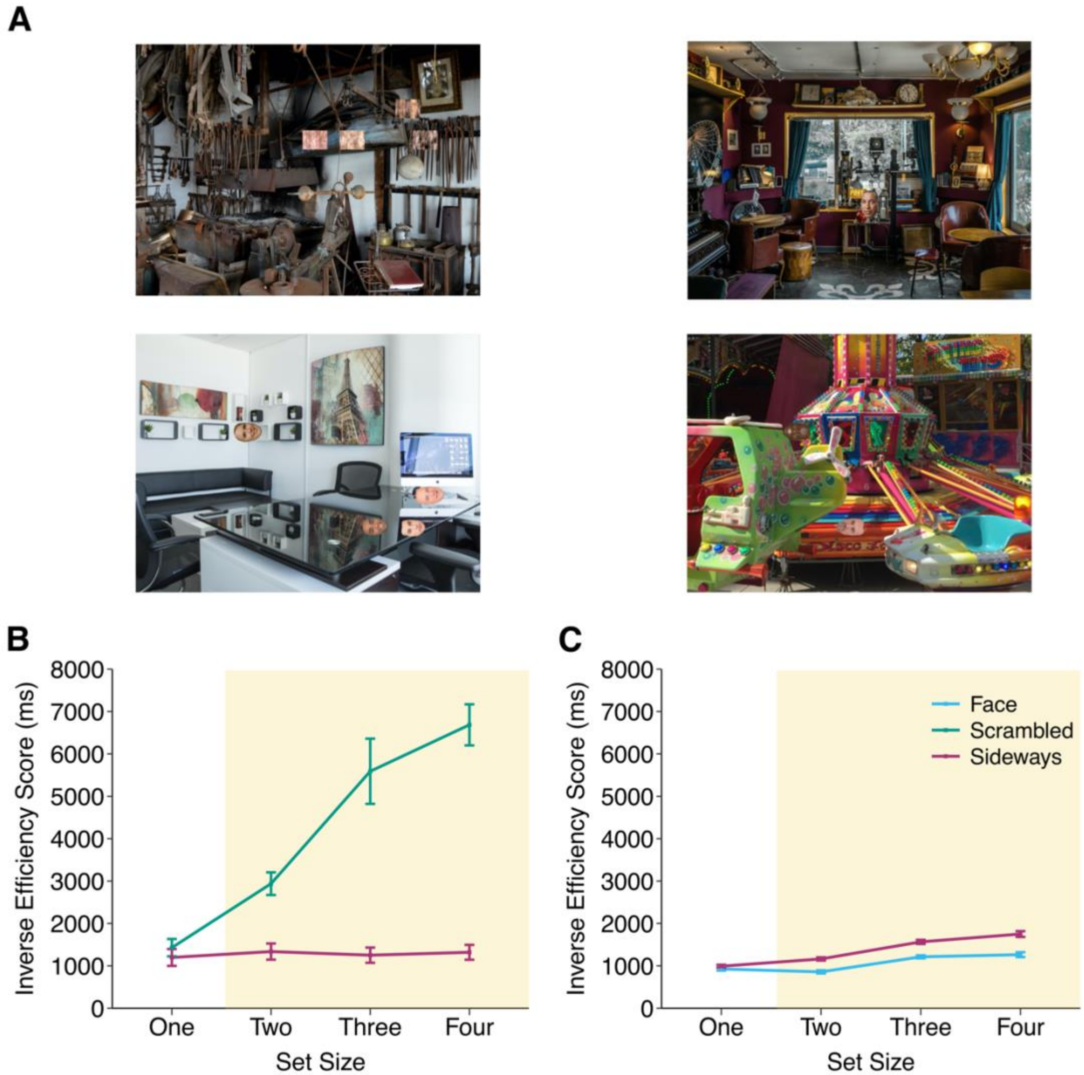

Figure 4.3 Example displays used in Experiments 12 (A – Left) and 13 (A – Right). Mean IES as a function of Set Size, shown separately for Sideways and Scrambled targets for Experiment 12 (B), and for Upright and Sideways targets in Experiment 13 (C). Lower scores indicate better efficiency. Error bars show within-subjects standard error (Cousineau, 2005). Yellow shaded area indicates a significant difference Sideways targets and Scrambled targets (Experiment 12), and Sideways targets and Sideways targets (Experiment 13), at a particular set size.

The results of Experiment 12 show overall higher detection efficiency for sideways faces compared to scrambled faces. Sideways faces incurred no detection cost across Set Size supporting a parallel detection process for up to 4 items tested here. These findings suggest that the face detection template is invariant to orientation. Moreover, the findings suggest that a pair of eyes – regardless of whether they are oriented horizontally or vertically – can facilitate efficient detection. Experiment 13 explores this further by comparing *Sideways Faces* to *Upright Faces*.

4.8 Experiment 13: Faces and Sideways in Real Scenes (250 ms)

Experiment 13 builds on Experiment 12 to investigate the specificity of the face detection template in multiple face detection. It directly compares *Upright Face* Targets with *Sideways Face* Targets to test the extent of the orientation invariance of the face template. If multiple face detection can proceed regardless of orientation, then both conditions should be detected with similarly high efficiency. However, if multiple face detection is specific to upright-oriented faces, then *Upright* Targets should be detected more efficiently. Based on the prior experiments, no cost per additional item is expected as Set Size increases for both *Upright* and *Sideways* Targets.

4.8.1 Methods

4.8.1.1 Participants

Seventy-one participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment. Eleven participants were excluded due to failed attention checks (2 or more within a block, or 3 across the entire experiment) or slow responses (> 2.5 SD from the group mean). The final sample (N = 60) comprised 37 females and 23 males (age range 19 – 42; M = 26.71, SD = 5.91).

4.8.1.2 Design and Stimuli

The design and scene stimuli were identical to Experiment 8. In this experiment, to-be-detected stimuli were *Upright Faces* Targets and the same

Sideways Face Targets from Experiment 12. Example displays for Experiment 13 can be seen in Figure 4.3 (A).

4.8.1.3 Procedure

The procedure in Experiment 13 is identical to Experiment 8 with displays presented for 250 ms.

4.8.2 Results and Discussion

Trials with a reaction time below 150 ms and above 3000 ms were excluded from the final analysis (< 0.01% of all trials). IES data was calculated using accuracy and reaction time. Separate analyses of both measures are provided in the Supplementary Materials and support the same conclusions as IES.

The separate 90° left and 90° right oriented face conditions were collapsed into a single *Sideways Faces* condition.

Figure 4.3 (C) shows IES data for each condition for Experiment 13. IES data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright*, *Sideways*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 34.71, $p < .001$, $\eta^2 = 0.37$]. Detection efficiency was greatest for *One* item (M = 959 ms, SE = 36 ms), followed by *Two* items (M = 1010 ms, SE = 37 ms), and then *Three* items (M = 1389 ms, SE = 42 ms), and for *Four* items (M = 1507 ms, SE = 63 ms). A significant main effect of Target Type was also found, with greater detection efficiency for *Upright* Targets (M = 1065 ms, SE = 42 ms) compared to *Sideways* Targets (M = 1368 ms, SE = 47 ms), [F (1, 59) =121.11, $p < .001$, $\eta^2 = 0.67$]. A significant interaction effect between Set Size and Target Type was also found [F (3, 177) = 17.99, $p < .001$, $\eta^2 = 0.23$].

Simple main effects revealed significant differences in detection efficiency across Set Size for *Upright* Targets [F (3, 354) = 16.04, $p < .001$, $\eta^2 = 0.12$]. One (M = 966 ms, SE = 38 ms) and Two (M = 858 ms, SE = 37 ms) Upright Targets were detected with similarly greater efficiency than Three (M = 1213 ms, SE = 42 ms) and Four (M = 1264 ms, SE = 63 ms) Upright Targets, which were detected with similarly poorer efficiency. The same pattern was observed for *Sideways* Targets [F (3, 354) = 47.82, $p < .001$, $\eta^2 = 0.29$, One Sideways, M = 991 ms, SE = 35 ms; Two Sideways, M = 1163 ms, SE = 38 ms; Three Sideways, M = 1566 ms, SE = 48 ms; Four Sideways, M = 1751 ms, SE = 68 ms].

Simple main effects between Target Type at each Set Size revealed no differences at Set Size One [F (1, 236) = 2.11, $p = .147$, $\eta^2 = 0.01$]. However efficiency for the *Upright* condition was significantly better than *Sideways* condition at each following Set Size [Set Size Two, F (1, 236) = 45.58, $p < .001$, $\eta^2 = 0.16$; Set Size Three, F (1, 236) = 60.94, $p < .001$, $\eta^2 = 0.21$; Set Size Four [F (1, 236) = 115.99, $p < .001$, $\eta^2 = 0.33$].

Experiment 13 shows overall greater detection efficiency for upright faces over sideways faces. These results suggest a preference in the face detection template for upright-oriented faces during multiple face detection. However, the detection of sideways faces was still efficient, suggesting that the face detection template can still detect these stimuli as faces regardless of orientation. The results of Experiments 13 and 12 support a detection template invariant to orientation.

Across Experiments 8 – 13, we investigated multiple face detection in real scenes and under different background manipulations. Regardless of background complexity and meaningfulness, face detection seems to operate in a highly efficient parallel manner. However, these experiments do not provide a complete picture of spontaneous multiple face detection. First, by giving specific instructions before the task, they do not adequately measure face detection as it might occur in the real world. Second, while the stimuli are highly controlled, they are not realistic. The faces cropped to outline lack the necessary body context that would be expected in the real world. Third, the stimuli and displays do not account for the sheer variability of the real world, which contains massive variations in lighting, colour, viewpoint, brightness etc. Fourth, a maximum of four targets were used in these experiments, whereas we are likely to encounter more than four faces at times. These various experimental design choices were helpful in the initial investigation of multiple face detection in real scenes. However, a move towards realistic images of real-world situations is necessary to fully understand how more than one face is detected. Experiment 14 address this.

4.9 Experiment 14: One–Shot Experiment

In Experiment 14, we aim to investigate spontaneous multiple face detection in naturalistic settings using hundreds of real images. To circumvent the limitations of previous experiments, we adopt a novel single-trial design whereby each participant is presented with a single unique image without repetition. First, only after image presentation are participants given the task instructions to report how many faces they saw. Second, by using real images of real people the stimuli retain body context and information that were removed when faces are cropped to outline in the previous experiments. Third, visual properties, such as brightness, contrast, pose, lighting etc, were allowed to vary as much as possible. Having context and variability in the stimuli mimics what we experience in real world situations. Fourth, only the number of faces was controlled, and Set Size could range from 0 – 6, while the number of possible responses was 0 – 9. Expanding both the number of faces and the range of responses allows us to investigate the capacity limit of multiple face detection by assessing if participants incorrect responses are under- or over-estimations.

The single-trial design is useful as it attempts to mimic spontaneous real-world face detection while also providing the means of assessing it in laboratory settings. To understand if we can detect faces even in these highly realistic examples, we would expect relatively high accuracy when faces are absent and when they are present, i.e., Set Size 0 and Set Size 1, respectively. If this is the case, then it would replicate previous findings of absent *vs* present face detection experiments. Moreover, if more than one face can be detected at a time, we would expect high accuracy past Set Size 1 up to Set Size 4, as tested in the previous experiments. Accuracy at larger Set Sizes (4 – 6) is expected to decrease as Set Size increases, but still remain above chance. Whilst incorrect responses at these larger Set Sizes are expected to be under-estimations of the actual number of faces as opposed to over estimations. If accuracy remains above chance, and incorrect responses are under-estimations, then this would point towards a detection mechanism that can detect a larger number of faces but with greater difficulty and poorer efficiency.

4.9.1 Methods

4.9.1.1 Participants

One thousand four hundred participants were recruited through Prolific recruitment service (www.prolific.co) and completed the experiment in exchange for a small payment. Twelve participants were excluded due to slow responses (>2.5 SD from the group mean). The final sample (N = 1388) comprised 652 females, 726 males, and 10 who preferred not to answer (age range 18 – 76; M = 29.89 SD = 10.51).

4.9.1.2 Design and Stimuli

Stimuli were gathered from *creative commons*, public Instagram accounts, and across the internet. A total of 700 naturalistic images were collected divided into 100 unique examples for each of the seven conditions of *Zero, One, Two, Three, Four, Five,* or *Six* people. Only images with full frontal, or partial frontal face view were selected. All other image properties including size, lighting; brightness; pose etc were left to vary.

A between-subjects design with the factor of Set Size (*Zero, One, Two, Three, Four, Five,* or *Six*) was used. Each participant saw a single unique image only once. Example images can be seen in Figure 4.4.

Figure 4.4 Example images from each set size used in Experiment 14. Zero faces (MacEntee, 2010); One Face (Ming, 2014), Two Faces (Almazawi, 2013); Three Faces (Ashleigh, 2012), Four Faces (Kamaludin, 2011), Five Faces (Nojiri, 2011), Six Faces (Infusionsoft, 2009).

The experiment was hosted online at Gorilla Experiment Builder (Gorilla.sc; Anwyl-Irvine et al., 2020). Participants could access the experiment on any desktop or laptop computer resulting in variable screen sizes. Mobile devices and tablets were excluded.

4.9.1.3 Procedure

The experiment was conducted in two identical runs of 700 unique trials. Each run consisted of 100 trials from each of the 7 Set Sizes.

Participants were asked to pay close attention to the screen but were not given specific instructions regarding the task. Once the participants indicated they were ready to begin the experiment a fixation cross appeared for 1400 ms with a 400 ms pause before and after. The stimulus image was then flashed for 200ms. Next, the participants saw a response screen where they were instructed to use either the number keys or the number buttons on screen to indicate how many faces they saw. Possible response options ranged from 0 – 9.

4.9.2 Results and Discussion

Total correct and incorrect participant responses were collated. Figure 4.5 shows the percentage distribution of each 0 – 9 responses given for each Set Size, as well as the percentage agreement between the paired displays in 1st and 2nd run of the experiment.

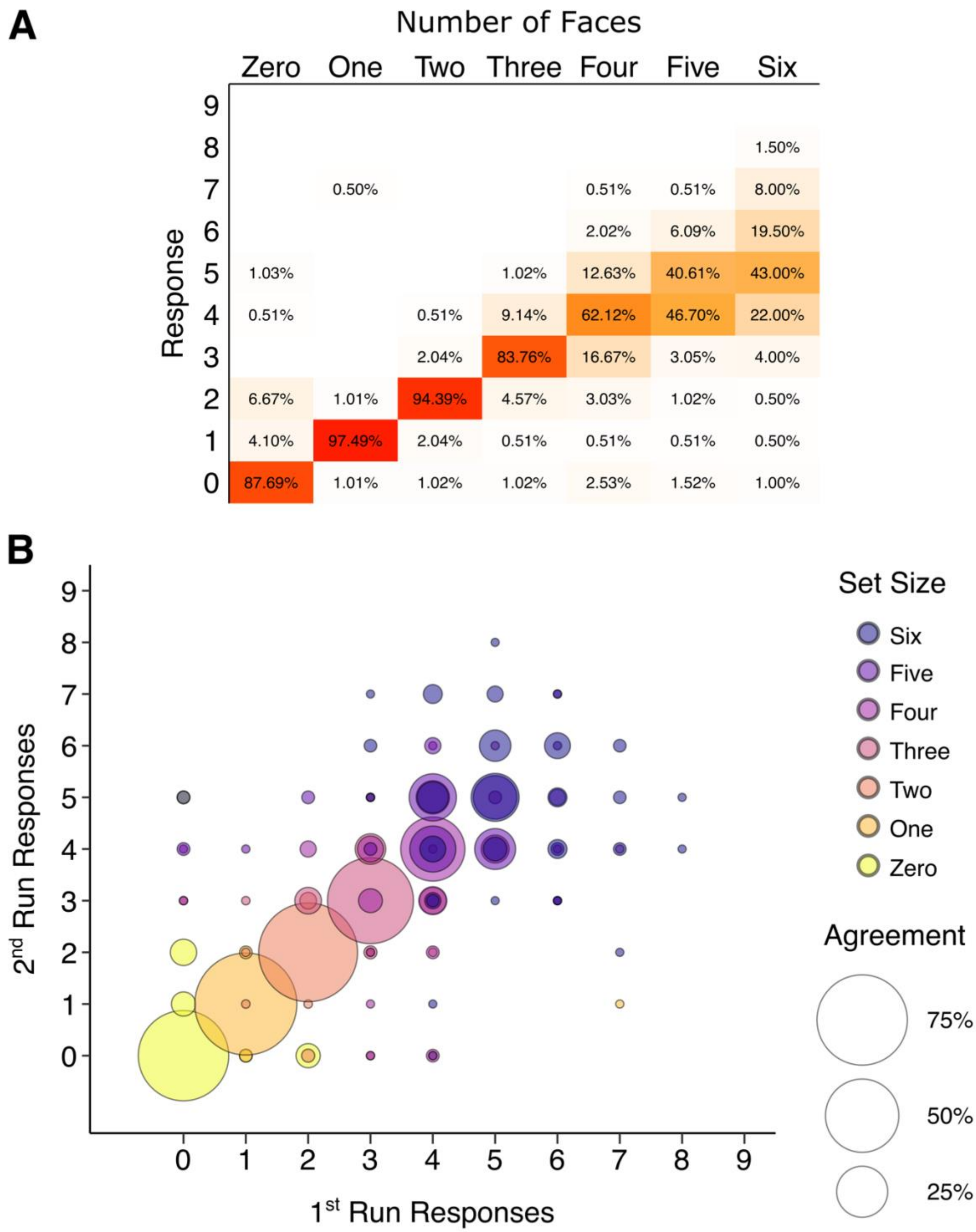

Figure 4.5 Heatmap (A) showing the percentage distribution of each response for each set size. Cells across the diagonal indicate percentage accuracy distribution. Cells off the diagonal indicate percentage error distribution.Bubble plot (B) showing percentage agreement between participants for paired displays in 1st and 2nd runs for each possible response value. Larger bubble size indicates greater agreement.

A Chi-Squared Test of Independence found significant differences in accuracy across Set Size, χ^2 (6) = 496.34, $p < .001$. Accuracy was high for when no faces were present in the scene at Set Size Zero (87.69%). When faces were present in the scenes accuracy was also high but only for the smaller Set Sizes of One (97.49%), Two (94.39%), and Three (83.76%). Accuracy then decreased by ~21% between each of Set Sizes Four (62.12%), Five (40.61%), and Six (19.50%).

A second Chi-Squared Test of Independence was conducted to investigate detection accuracy between the smaller Set Sizes (One, Two and Three) and larger Set Sizes (Four, Five, and Six). Significantly higher detection accuracy was found for smaller Set Sizes (M = 91.88%, SD = 7.20%), compared to larger Set Sizes (M = 40.74%, SD = 21.31%), [χ^2 (1) = 350.26, $p < .001$].

To investigate whether participants over- or under-estimate the number of faces in a scene, errors across Set Sizes 1 – 6 were analysed with a further Chi-Square Test of Independence. A significant difference in the type of errors was found, χ^2 (5) = 69.28, $p < .001$. Overall errors were low at the smaller Set Sizes 1 – 3, and equally likely to be over- or under-estimations. At the larger Set Sizes, errors were more likely to be underestimations of the actual number of faces (Set Size Four, 60.00%; Five, 88.88%; Six, 77.02%).

The high accuracy at Set Sizes Zero and One suggests that we can differentiate between the presence and absence of a face in a real scene. But can we detect more than one face at a time? The high accuracy at Set Sizes Two and Three suggest that we can detect more than one face at a time from real scenes. Past these smaller Set Sizes, detection accuracy decreases incrementally but remains above chance (>10%). Furthermore, as can be seen in Figure 4.5(A), incorrect responses at Set Size Five and Six are concentrated below the diagonal, pointing toward a tendency to underestimate the number of faces. The agreement bubble plot in Figure 4.4(B), further supports this finding with higher agreement at small Set Sizes and declined at larger Set Sizes. Participants are still able to detect more than three or four faces, however this is a more difficult and inefficient process. Experiment 14 goes beyond the previous experiments and shows that spontaneous real-world multiple face detection can and does occur in a parallel process.

4.10 General Discussion

In a series of seven experiments, we asked participants to report the number of target items presented in visual displays. When these targets were intact upright faces (Figure 4.6), we saw no change in detection efficiency as the number of targets increased from one to four. The only exceptions were Experiment 10 (when targets were presented against blank backgrounds) and Experiment 13 (when vertical symmetry was broken). We conclude that viewers can detect multiple faces simultaneously.

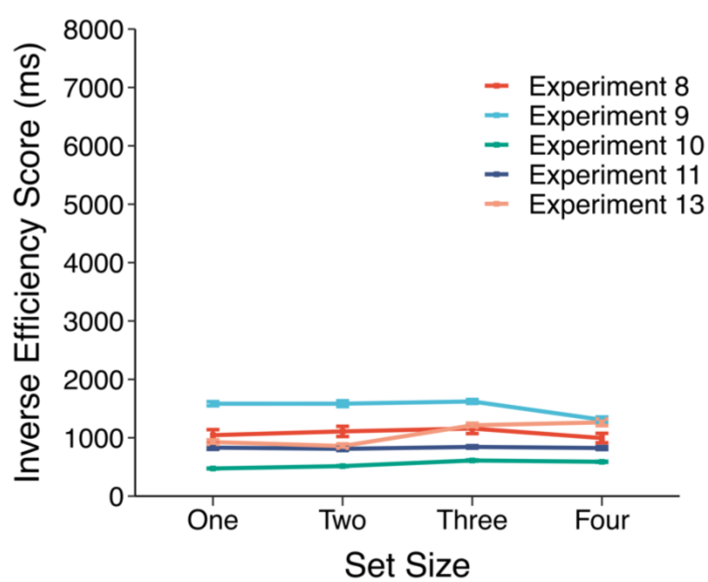

Figure 4.6 Summary detection efficiency (IES) as a function of set size for upright faces in Experiments 8–11 and 13. Error bars show within-subjects SE (Cousineau, 2005).

This conclusion underscores the key findings of Qarooni et al. (2022) and generalises them beyond psychophysical methods. First, the current studies demonstrate multiple face detection in the context of a search task, in which spatial uncertainty requires localisation of targets as well as categorisation (Bindemann & Lewis, 2013b). The demand to localise was complicated by visual clutter—naturalistic scenes in Experiments 8, 9, 10, 11, and 13; Voronoi scenes in Experiment 11. Unlike blank backgrounds (Experiment 10), cluttered scenes require the viewer to segment face-containing regions from competing regions that are visually busy. Our results show that viewers can make the necessary discrimination at several locations simultaneously.

Second, we estimate an upper bound on multiple face detection. Beyond four faces (Experiment 14), performance became markedly less efficient. Previous experiments have found equivalent face detection efficiency over the range 2–4 faces (Qarooni et al., 2022). However, while those experiments showed that capacity limits in face detection can exceed one face, they did not test for an upper bound; nor did they allow direct comparison of single versus multiple face detection.

Third, we examined the effects of several visual factors on performance, delineating conditions under which multiple target detection is likely to be efficient or inefficient. In the following sections, we consider each of these factors in turn.

4.10.1.1 Effects of exposure duration

In Experiment 8, displays were presented for 250 ms, whereas in Experiment 9, they remained on screen until response. In both of these experiments, detection was similarly efficient for upright and inverted faces. The small numerical advantage for upright faces was not statistically significant. Critically, there was no cost of increasing set size over the tested range of (1–4 items), indicating rapid (<250 ms) parallel detection for upright and inverted faces alike.

Detection was much less efficient for scrambled faces. In Experiment 8 (limited exposure) increasing the number of scrambled faces impaired performance. This observation confirms the effectiveness of our set size manipulation, and implies serial detection for scrambled faces in this context. In Experiment 9 (unlimited exposure), participants could use additional viewing time to improve their performance. Even so, efficiency for scrambled faces did not approach efficiency for intact faces at any set size. In sum, restricted viewing time did not present a problem for intact faces; unrestricted viewing time did not solve the problem for scrambled faces.

4.10.1.2 Effects of visual background

Search tasks naturally incline us to focus on search targets. But comparison of Experiment 8 (cluttered scenes), Experiment 10 (blank background), and Experiment 11 (Voronoi scenes) underscores the importance of visual context in determining task performance.

Compared with a cluttered scene, a blank background greatly improved overall efficiency in this task. It also eliminated the main effect of the target type. This is perhaps unsurprising, given the high salience of any of these targets against the uniform grey surround. More interesting is that the blank background was associated with a small but significant effect of set size for all target types, possibly reflecting spatial uncertainty; cf. Qarooni et al. 2022).

This pattern of findings for blank backgrounds is evidently fragile. Merely changing the background to a meaningless Voronoi scene (or a meaningful cluttered scene) was enough to abolish the effect of set size on upright faces, the overall high efficiency, and the convergence of target types. The fragility of the blank background findings urges caution when generalising to other viewing conditions. For non-blank backgrounds, we repeatedly found that viewers could detect multiple faces simultaneously.

4.10.1.3 Effects of target type

With the exception of Experiment 10 (blank backgrounds), scrambled faces were detected less efficiently than upright or inverted faces. Given that all three target types were matched in terms of overall size, shape, colour, and low-level visual energies, the scrambled face deficit implies that the spatial structure of these energies is critical for detecting faces in scenes. This interpretation echoes previous studies (Purcell & Stewart, 1986, 1988). What the current experiments reveal is how profoundly set size can exacerbate performance differences (e.g. Experiments 8 and 12). If we had only presented single targets in these experiments, we would have missed the sharp divergence between scrambled and intact faces. Only by testing multiple target detection can we see this bigger picture.

Across experiments, detection efficiency for upright and inverted faces was virtually indistinguishable—a finding that resonates with visual search for faces (Bindemann & Burton, 2008; Lewis & Edmonds, 2005). This observation raises the question of category membership. What do upright and inverted faces have in common that scrambled faces do not? One possibility that has clear theoretical precedent is vertical symmetry. For example, the idea of a 1D ‘bar code’ for face perception relies on vertical symmetry (Dakin & Watt, 2009b)A related possibility is horizontally aligned eyes. A pair of eyes is a potent stimulus for the human attention system (Kingstone et al., 2004), and eyespot mimicry throughout the animal kingdom suggests that the salience of horizontal eye pairs is deeply rooted (Radford et al., 2020; Skelhorn et al., 2016)

One aspect of our findings suggests a role for such factors in multiple face detection. Although inverted faces were detected just as efficiently as upright faces (Experiments 8–11), sideways faces were detected less efficiently (Experiment 13). As this divergence between upright and sideways faces only emerged from set size two, it could only be revealed by presenting multiple targets.

The findings of Experiments 8–13 can be summed up as follows: (i) Viewers could detect several faces at once, even in cluttered scenes. (ii) Efficient performance was contingent on the internal spatial layout of the face, not overall colour or outline. (iii) Multiple face detection was inversion invariant, but not orientation invariant.

4.10.1.4 Spontaneous multiple face detection

Experiment 14 represents a deliberate departure from the preceding series. Using entirely different stimuli and methods, we found that viewers could detect three or four faces concurrently, even when single-trial presentation and retrospective task assignment ruled out expectation or practice effects. We also showed that multiple face detection generalises to coherent human figures photographed in hundreds of naturalistic scenes.

Presenting up to six faces allowed us to estimate an upper bound for capacity limits in face detection. From set size five, errors outnumbered correct enumeration responses. Providing an expanded range of response options (zero to nine) allowed us to distinguish overestimates from underestimates at all set sizes. Underestimates were the dominant error from set size four. We propose that, under conditions of spatial uncertainty and in cluttered scenes, the upper limit for multiple face detection is four, plus or minus one.

Our findings suggest a number of avenues for future research. In particular, it will be interesting to estimate the separate contributions of perceivers, scenes, and facial appearance to overall variability in face detection span. Given the importance of faces in social cognition and its disorders, an individual differences approach seems especially promising (Eayrs & Lavie, 2021b). For example, measuring the association between face detection span and social anxiety (Doty et al., 2013) or ASD (Schauder et al., 2019) would provide a novel test of plausible etiological pathways. Individual differences in face detection span could be assessed by combining the multi-trial design of Experiments 8–13 with the expanded response options of Experiment 14.

Another interesting question concerns the role of top-down expectations in multiple face detection. We saw in Experiment 14 that underestimates were more frequent than overestimates past four faces (Figure 4.5 heat map). One possibility that could account for this is social knowledge. Some human group sizes are more common than others, notably groups in the 3–5 range (Dunbar & Spoors, 1995; Zhou et al., 2005). Presumably, these regularities become internalised by members of society—either through the statistics of everyday experience or over the course of human evolution. Such top-down knowledge could influence perception when the bottom-up signal is weak (e.g. brief presentation and high set size in Experiment 14), thus biasing estimates towards the expected range. Future experiments could measure top-down influence more directly by manipulating perceptual uncertainty. For now, we show that humans can detect up to four faces (plus or minus one) in complex scenes efficiently and concurrently. Our findings illuminate an early process in social cognition that is often neglected. They also show that presenting multiple targets can transform our understanding of search performance.

Chapter 5 – General Discussion

The current thesis had three main aims. The first aim was to compare multiple target detection for faces and other types of stimuli. The second aim was to estimate capacity limits in face detection and investigate the serial-*vs*-parallel nature of face template matching. The third aim was to test how multiple face detection is affected by different viewing conditions, including aspects of the task and presentations. Achieving these three aims improves our theoretical understanding of the quantitative dimension of face detection. It also expands our knowledge of the qualitative dimension. Moreover, the novel paradigms developed in this thesis contribute several new experimental methods for assessing face detection specifically.

This general discussion shall briefly summarise the main findings of each experimental chapter. It shall then discuss the theoretical implications and methodological contributions of these findings. Finally, it shall reflect on the remaining questions of the capacity limits of face detection and suggest future directions for investigating these questions.

5.1 Overview of the Current Experiment Work

The three experimental studies within Chapter 2 aimed to distinguish between multiple face detection and multiple non-face detection. Using a 'subitizing of faces' approach, two main findings were established. First, detection of up to four faces outperformed detection of up to four non-faces. Second, upright and inverted faces were indistinguishable from each other based on detection performance. Together these findings establish a foundation for the subsequent studies in this thesis. More specifically, they point towards a special detection mechanism for faces over non-faces that can be sustained for up to four faces.

Chapter 3 focused on addressing the capacity limits of face detection to assess the serial-*vs*-parallel nature of the process. The novel 'fixed/mixed' judgement task allowed for an estimate of the cost per additional item in the display without relying on visual search or enumeration. The four experiments reported in Chapter 3 revealed a parallel detection process in which two, three, or four faces could be detected with no discernible cost per item. However, this parallel cost-free detection was contingent on the surrounding visual context. Supporting prior findings from Chapter 2, Chapter 3 also found a serial detection process for upright and inverted faces, which could not be distinguished from each other. Overall, these results further support an inversion-invariant detection mechanism. They also aid in locating the bottleneck in face perception.

The final series of experiments reported in Chapter 4 investigated multiple face detection from real scenes and factors that may affect it. First, even from real photographs of complex scenes, face detection proceeds in an efficient parallel manner for up to four faces. Beyond this range, enumeration accuracy declined, and underestimates were more likely than overestimates. Second, incorporating visual context in face detection experiments is vital to assess the process as it would occur in everyday life. Third, a small detection advantage was seen for upright faces in real scenes compared to sideways faces. Together these results demonstrate how face detection is a process that is largely unaffected by the complexity of the visual background or by inversion of the face. They also emphasise the importance of both the 'template' and the 'visual environment' aspects of face detection.

5.2 Theoretical Implications

The experimental chapters in this thesis suggest that detecting multiple faces from complex visual backgrounds is a parallel and efficient process for up to four faces. This section discusses the theoretical implications of these findings in the context of the current face detection literature. Specifically, it addresses quantitative aspects of face detection by discussing how we can detect multiple faces at once and how this finding adds to our understanding of the bottleneck of face processing. Next, it addresses qualitative aspects of face detection, including the observed inversion-invariance and the issue of category membership (i.e. what counts as a face). Finally, it will consider social and evolutionary implications of the ability to detect multiple faces.

5.2.1 Quantitative Aspects of Face Detection

This thesis provides two main takeaways regarding the quantitative aspect of face detection. First, we can detect multiple faces at once. This finding stands in contrast to later face processes, which appear to be strictly limited to one face at a time. Second, we can detect up to four of these faces in an efficient and parallel manner. These findings intersect with both numerical cognition and face detection literature.

Face detection and subitizing share similar capacity limits and parallel processing of up to four items suggesting they may also share similar underlying mechanisms. To account for our subitizing abilities, Trick & Pylyshyn (1994) proposed the FINSTs theory. This theory suggests the presence of a capacity-limited process of item individuation. Up to four mental indices – termed FINSTs – can be occupied parallelly to track objects in the visual environment. Both subitizing and the FINSTs theory are domain-general in the sense that any object, regardless of its properties – or even if these properties change – can be enumerated and tracked in the visual environment (Katzin et al., 2019; Pylyshyn, 2004; Trick, 1992). Under this domain-general perspective, both faces and non-faces should be detected in parallel up to four items. Indeed Experiments 1 and 2 in this thesis show that faces and non-faces alike can be detected/subitized accurately and rapidly. However, across all experiments in this thesis, including Experiments 1 and 2, face detection outperforms non-face detection. This finding supports the face detection advantage noted in previous face detection work (Crouzet et al., 2010; Keys et al., 2021; Purcell & Stewart, 1986, 1988; Wardle et al., 2020). This general discussion argues that even though subitizing and, in turn, FINSTs theory are domain-general, there appears to be priority for detecting multiple faces over non-face stimuli. In the context of the FINSTs theory, it may be that priority is given to faces over non-faces to occupy the mental indices. This view of a 'special' status for faces at the detection stage fits well within established literature on the 'special' status for faces in later processes (Farah et al., 1998; Kanwisher et al., 1997; Tanaka & Farah, 2007).

The finding of a parallel cost-free detection mechanism for up to four faces also resolves some discrepancies in the face detection literature. This specific detection pattern seems unaffected by some manipulations of the visual environment and emerges more clearly when faces are in real scenes. However, when faces are embedded amongst inverted faces (i.e. similar stimuli), detection shifts to a serial process where each additional face incurs a cost. Serial processing has been previously suggested by Nothdurft (1993), who found that search times increased as the number of jumbled or inverted face distractors increased. In this case, the target and distractor were similar. In contrast, Lewis & Edmonds (2005) argued for parallel processing after finding no increase in search times for a single target upright face as the distractor scrambled scene size increased. But in Experiment 2 of that study, Lewis & Edmonds (2005) also found serial detection when faces are embedded amongst inverted faces. The findings in this thesis suggest that the discrepancies in previous findings may be due to the high level of similarity between upright and inverted faces. The notion of inversion-invariance is examined in section 5.2.2. For now, inversion-invariance does seem to account for serial detection pattern found whenever upright and inverted faces must be differentiated.

Nonetheless, inversion-invariance also supports the staged detection mechanism proposed by Lewis and Edmonds (2005). It appears that an initial stage relies on the broad tuning and orientation-invariance of the face detection template to select any plausible candidate face region, which could include upright faces, inverted faces, sideways faces or pareidolic faces (Wardle et al., 2020). Then a second, more selective stage deals with orientation (if necessary) followed by later face processes. This staged detection process would account for the serial and costly detection displays that combined upright and inverted faces in this thesis. If all objects in these displays are detected as faces, then the detection process must slow down at the second stage to assess each face and rotate it if needed.

The observation that we can detect multiple faces at once has also aided in locating the bottleneck in face perception. Later face processes, such as extraction of personal identity or semantic information, appear to be strictly capacity limited to one face at a time (Bindemann et al., 2007; Bindemann, Burton, et al., 2005). In contrast, face detection appears to have a capacity limit of up to four faces. Two alternative hypotheses regarding the bottleneck of face processing were outlined in the General Introduction and are depicted in Figure 1.3. The observed capacity limit of later face processes could originate at the detection stage, such that detection too is capacity limited to one face and constrains later processes. Alternatively, capacity limits could originate only after the detection stage, such that the detection process can accommodate more than one face, but later processes cannot. As discussed in Chapter 3, the current findings point to a bottleneck in face processing *after* the detection stage, implying different processing bandwidth for detection and later face processes. Figure 5.1 illustrates this revised hypothesis regarding the bottleneck of face processing.

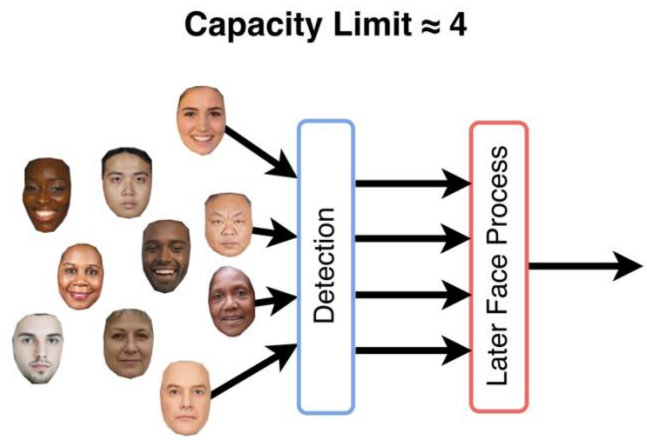

Figure 5.1 Revised hypothesis for the bottleneck in face perception. Up to four faces can be detected in parallel suggesting the bottleneck is present *after* the detection stage.

It is important to note that while up to four faces can be detected at once, capacity limits of face detection may still be malleable to some extent. Individual differences in processing capacities, visual load, and attention have been found to affect visual item enumeration (Alvarez & Cavanagh, 2004; Cavanagh & Alvarez, 2005; Cowan, 2001; Eayrs & Lavie, 2018; Eayrs & Lavie, 2021; Railo et al., 2008). Moreover, accuracy results from Experiment 14 further suggest that capacity limits can fluctuate across specific displays. Underestimations were more likely than overestimations when five or six faces were presented in photographs. But accuracy at these set sizes was still above chance, suggesting that in some instances, five or six faces may be detected. The one-participant-one-trial design of Experiment 14 did not allow for further explorations of expanded capacity limits, be they individual differences, scene properties, or clustering of individual faces. However, the groupitizing phenomenon from numerical cognition, in which subitizing of items is facilitated through grouping, may offer one explanation. This point is returned to in section 5.4, which discusses avenues for future research. Nonetheless, whether or not capacity limits can expand or contract under certain special circumstances, it appears that a face detection span of four (plus or minus one) is typical for everyday face detection.

5.2.2 Qualitative Aspects of Face Detection

While this thesis has focused on quantitative aspects of face detection, qualitative aspects cannot be separated entirely, as multiple face detection requires multiple template matching. Throughout the experimental work presented here, whenever the face detection template is matched to different regions of the visual environment, interesting findings emerge regarding the different categories of objects. Specifically, there appears to be an inversion-invariant aspect to the detection template such that upright and inverted faces are detected in a similar manner. In the context of previous face detection literature, this points to a hierarchy in the detection of different classes of objects.

The notion of an inversion-invariant face detection template contrasts with the face detection effect (FDE) reported by Purcell & Stewart (1986, 1988). The FDE found shorter presentation times were required for intact upright faces compared to inverted or jumbled faces. Inversion-invariance at the detection stage also contrasts with previous face perception literature that has long touted faces as 'special' stimuli preferentially processed over inverted faces and non-faces (Farah et al., 1998; Kanwisher et al., 1997; Tanaka & Farah, 2007; Yin, 1969b). In the context of multiple face detection, this inversion-invariance seems to come into effect in the earliest stages of detection, as suggested by Lewis and Edmonds (2005). It also suggests broad tuning in the face detection template, such that the template is selective for face stimuli but not especially sensitive to their vertical orientation (i.e. upright or inverted). This finding echoes recent face detection work exploring similar attentional bias to upright and inverted faces (Bindemann & Burton, 2008; Devue et al., 2012; Laidlaw et al., 2015). Moreover, this account could also explain why sideways faces in Experiment 13 were detected efficiently, but not as efficiently as upright faces.

The inversion-invariance of the detection template appears to hinge on a general abstract representation of faces which may contain common features shared between upright and inverted faces – as suggested by Bindemann & Burton (2008) and Lewis and Edmonds (2005). Horizontal eye pairs (Kauffmann et al., 2021; Omer et al., 2019) and horizontal spatial frequency 'bar codes' within the face (Dakin & Watt, 2009a; Goffaux & Dakin, 2010) have both been proposed as components of the detection template. Both these visual properties are preserved when a face is upright or inverted, but not when it is scrambled. Yet the major divergence between sideways faces and scrambled faces (Experiment 12) shows that eliminating these properties did not reduce a sideways face to 'non-face' status in this task. Conversely, the minor divergence between sideways faces and upright faces (Experiment 13) suggests a possible role for these properties in mediating match to template. It is outside the scope of this thesis to determine the exact properties of the face detection template. But future work could adapt the methods developed in this thesis to pursue these qualitative questions. For instance, the fixed/mixed design could be adapted to compare upright faces with comparison stimuli that have been modified in other ways (e.g. orientation of eye pairs).

Orientation aside, there does appear to be a detection advantage for faces as a category, relative to other classes of objects. Previous face detection eye-tracking and visual search studies have found preferential detection of faces over face-like stimuli (pareidolic and illusionary faces), which in turn, are preferentially detected over non-faces (Crouzet et al., 2010; Keys et al., 2021; Wardle et al., 2020). Interestingly, the nuances of this hierarchy point to a detection template that is selective for faces, but broadly tuned. Wardle et al. (2020) investigated the brain representation of faces, pareidolic faces, and matched non-face objects in an fMRI and MEG study. The representation of pareidolic faces, i.e. illusionary faces, was initially more similar to real faces rather than to matched non-face objects. However, within 250 ms, this representation changed to be more like non-face objects than faces. The similar neural representations for faces and pareidolic faces were also echoed in a follow-up visual search study, which found fastest search times for faces, followed by pareidolic faces, followed by non-face objects (Keys et al., 2021).

5.2.3 Detecting Multiple Social Stimuli

The detection hierarchy of different classes of objects seems to be topped by faces, a specific class of highly salient social stimuli. Moreover, the broad tuning of the detection template seems to suggest a specialised detection mechanism that errs on the side of caution, preferring false positives to misses. Regions of the visual environment that contain candidate faces are apparently detected in a rapid parallel manner. It is only after this initial detection stage that a more discerning process is implemented, perhaps through inversion rotation as suggested by Lewis and Edmonds (2005), or through changes in brain representation for pareidolic faces, as observed by Wardle et al. (2020) and Keys et al. (2021).

Detecting multiple faces in parallel may serve as an evolutionary advantage, allowing rapid identification of the potential threats and benefits in the visual environment. Supporting this claim, New et al. (2007) found preferential and category-specific attention to animate over inanimate objects, arguing that it was membership to "ancestrally important categories" (pg. 16598) that drove this attention. Furthermore, experiments have found highly efficient visual search for human faces amongst primate or mammal faces (Simpson, Buchin, et al., 2014; Simpson, Husband, et al., 2014). These findings again point to a hierarchy within detection. Even amongst other non-human faces, a preference still persists for faces. With humans being highly social animals, whose success as a species could be attributed to our ability to socialise and understand each other (Dunbar, 2012; Tomasello, 2019), an expanded ability to assess the presence of faces would confer an advantage.

5.3 Methodological Contributions

The finding of a parallel detection mechanism for up to four faces was made possible by adapting existing paradigms from numerical cognition and face perception to the current research questions. To date, most subitizing work has incorporated simple circle stimuli in displays with high control between luminance, surface area, size, and density (De Marco & Cutini, 2020), and only a few have incorporated other stimuli such as geometric shapes (Vuilleumier & Rafal, 2000), or human figures (Railo et al., 2016). At the same time, previous face perception research has relied on visual search or face categorisation tasks with a single face target (e.g. Lewis and Edmonds, 2005; Liu, Harris, & Kanwisher, 2002). The experimental work in this thesis has combined methods from both fields to assess multiple face detection.

The first methodological contribution is the simple manipulation of adding multiple faces to the display. Across this thesis, comparisons of different items at set size one often failed to distinguish between faces and non-faces. If detection was assessed based on the difference between one face compared to one non-face, then no detection advantage would be noticed. Only when multiple faces or non-faces were presented could divergent patterns of performance be discerned. While this methodological contribution is simple, it is impactful. To understand multiple face detection, we need to present multiple faces.

The second methodological contribution is the development of the fixed/mixed judgment task. Unlike visual search, the fixed/mixed task only requires participants to indicate whether or not all items in a set belong to the same category, without requiring them to know how many items are there and what type of items they are. This task directly assesses multiple face detection without directly measuring visual item enumeration of object category discrimination. It also lends itself to different manipulations. As mentioned earlier, qualitative aspects of the face detection template could be explored using the fixed/mixed task. For instance, detection of intact face stimuli could be compared to manipulated face stimuli to test whether the face detection template can (i) generalise to the manipulated faces, and (ii) similarly detect multiple faces and multiple manipulated faces. Affirmative results would provide converging evidence for a selective but broadly tuned detection template. Thus, the fixed/mixed task can be used to investigate both quantitative and qualitative aspects of face detection.

Third, it is important to consider the role of visual context. The pattern of a parallel face detection mechanism only emerged when multiple faces were embedded in complex visual backgrounds (e.g., Experiments 8 and 9). When this background was removed, parallel face detection was compromised (Experiments 10). Incorporating complex visual backgrounds into face detection experiments seems logical, if detection involves registering the presence of faces by comparing regions of the visual environment to a stored face template (Lewis & Ellis, 2003; Robertson et al., 2017; Tsao & Livingstone, 2008). If the visual environment is removed, then the number of candidate regions is drastically reduced, affecting detection. This situation could plausibly manifest as a rapid but serial process, as those remaining regions have to be assessed in turn to *categorise* them as faces rather than to localise them.

A further benefit of incorporating complex backgrounds is that they more closely resemble the visual environments in which everyday face detection occurs. However, everyday face detection is also spontaneous; we neither consciously prepare to detect faces nor instruct our face processing system to detect the faces we see. Visual search for faces departs from this spontaneous situation by prescribing an experimental task in which faces are defined as targets. The one-trial one-participant design of Experiment 14 hinges on giving participants the task instructions only *after* the display is presented. This experimental design appears to successfully tap into spontaneous multiple face detection. Participants could not prepare to report the number of faces they saw, but their detection system could efficiently detect up to four faces in parallel. By extending the range of possible responses beyond the number of presented faces, this experiment also allowed us to estimate the upper limit of face detection. Even though five or six faces can be detected with above-chance accuracy, viewers were more likely to generate underestimates at these large set sizes than to respond accurately or overestimate. More generally, Experiment 14 shows how a large-scale single-trial experiment can allow us to tackle research questions that do not fit conventional experimental designs.

5.4 Future directions

The theoretical and methodological advances presented here open up the neglected quantitative dimension of face detection. Their position at the intersection of face perception and numerical cognition show how these seemingly disparate areas can profit from each other. As well as addressing the research questions that first motivated this project, the thesis also identifies new research questions which can be explored in future research.

The later parts of this thesis examined how visual factors, including surrounding visual context, affect face detection span. The typical limit appears to be around four faces, plus or minus one, across the conditions tested here. However, future research could assess the prospects for extending this limit, and numerical cognition paradigms offer novel ways to do so. Groupitizing is a phenomenon within subitizing in which items that are chunked together are enumerated together (Anobile et al., 2020; Gilmore et al., 2018b; Starkey & McCandliss, 2014). One interesting question is whether groupitizing can apply to multiple face detection, such that the apparent capacity of four faces can be artificially inflated. For instance, eight faces in a display could be presented in four groups of two faces each. From a groupitizing perspective, all four pairs would be detected as individual units; however, they would be enumerated as eight items. If parallel face detection is strictly limited to four faces, in a way that is immune to grouping principles, then the results should replicate those seen in this thesis. However, if groupitizing applies to face stimuli as it applies to simple stimuli, then all eight faces should be detected at once.

Chunking of social stimuli seems plausible. For example, the social binding hypothesis (Figure 5.2) suggests that within scenes, social interaction between people binds them into groups (Vestner et al., 2019). Vestner et al. (2019) find support for a group-based analysis of social scenes whereby an interacting dyad is perceived as a single unit. Preferential detection for social pairs is also seen in the work of Railo et al. (2016). In their work, up to three human figures could be detected rapidly and accurately, but performance was best for two figures. Following the social binding hypothesis, if we have the capacity to detect up to four faces simultaneously, but socially relevant stimuli can be grouped together in pairs, it may be possible to push the limits of face detection beyond the capacity of four.

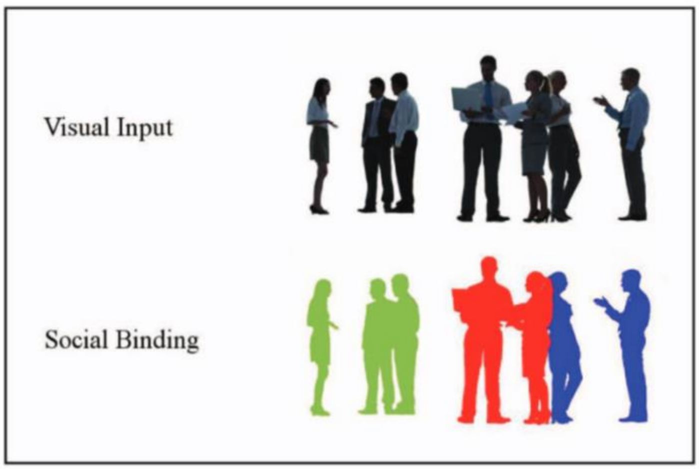

Figure 5.2 Visual representation of the Social Binding Hypothesis from Vestner et al. (2019). When viewing social scenes (top) it may be possible that interacting pairs are binded (bottom) together to simplify visual input and facilitate processing.

Future research into quantitative aspects of face detection can also help us understand the strengths and limitations of our own cognitive abilities, relative to those of computers. The vast majority of research into face detection concerns machine vision, not human vision. Computer algorithms are not restricted to the same capacity limits that biology imposes on humans. An increasingly pressing question is how humans can relate to cognitive abilities that greatly exceed our own—particularly in the social realm where we take our fluency for granted. It seems plausible that understanding our own cognitive limit will become important for supplanting intuitions of what machine systems can or cannot do, and calibrating expectations more accurately. Given the rate of progress in automated systems, face perception may be a useful domain in which to study this emerging dynamic in human-computer interaction.

5.5 Concluding Remarks

By volume, research on later stages of face perception greatly exceeds research on face detection. Even within face detection, research on qualitative aspects exceeds research on quantitative aspects. Using novel paradigms inspired by the numerical cognition literature, current work tackles several quantitative questions concerning face detection by humans. The thesis establishes that detecting multiple faces in complex visual scenes is an efficient parallel process for up to four faces. It also contributes methodological innovations that can be adapted to address related research questions. Face detection and later stages of face processing are governed by different perceptual principles.

Appendices

A.1 Supplementary Analyses: Chapter 2

This section contains the supplementary material for Chapter 2: Subitizing Faces, including full accuracy and reaction time analyses of Experiment 1 – 3.

A.1.1 Experiment 1: Absolute Subitizing 16 ms Exposure Time

A.1.1.1 Detection Accuracy

To investigate detection accuracy, a two-way ANOVA with repeated measures of Set Size (*One, Two*, *Three, Four, Five, Six, Seven, Eight*) and Item Type (*Upright Face, Inverted Face, Upright Non-Face, Inverted Non-Face*) was conducted. Table A.1 displays accuracy means and SE. The analysis revealed a significant main effect of Set Size, with accuracy decreasing as the Set Size increased [$F\ (7, 273) = 248.50$, $p < .001$, $\eta^2 = 0.86$], as well as a main effect of Item Type, with greater overall accuracy for *Upright* and *Inverted Faces* compared to *Upright* and *Inverted Non-Faces,* [$F\ (3, 117) = 4.13$, $p = .008$, $\eta^2 = 0.10$]. A significant interaction effect between Set Size and Item Type was also found [$F\ (21, 819) = 4.46$, $p < .001$, $\eta^2 = 0.10$].

Table A.1 *Experiment 1 mean % accuracy data for Upright Face, Inverted Face, Upright Non-Face, and Inverted Non-Face conditions at each Set Size. Within-subjects standard error in brackets (Cousineau, 2005).*

Set Size	Upright Faces	Inverted Faces	Upright Non-Faces	Inverted Non-Faces	Grand Mean
One	97.19 (0.96)	96.25 (1.00)	96.67 (0.95)	96.77 (0.99)	96.72 (0.97)
Two	93.65 (1.03)	93.02 (0.96)	88.96 (1.20)	88.65 (1.45)	91.07 (1.16)
Three	79.58 (1.98)	78.23 (2.13)	71.98 (2.03)	72.29 (1.92)	75.52 (2.02)
Four	70.21 (2.10)	68.23 (1.88)	60.94 (2.27)	60.73 (1.87)	65.03 (2.03)
Five	53.54 (2.53)	55.94 (2.39)	48.33 (1.86)	49.38 (2.61)	51.80 (2.35)
Six	47.40 (1.47)	42.50 (1.83)	47.50 (1.90)	45.83 (2.19)	45.81 (1.85)
Seven	39.17 (1.93)	39.17 (1.98)	42.92 (2.12)	42.81 (2.37)	41.02 (2.10)
Eight	26.25 (2.48)	26.35 (2.72)	32.08 (2.22)	32.60 (2.64)	29.32 (2.51)
Grand Mean	63.37 (1.81)	62.46 (1.86)	61.17 (1.82)	61.13 (2.01)	62.03 (1.87)

Simple main effects revealed significant differences in accuracy for *Upright Faces* as each Set Size increased [F (1, 1092) = 163.18, $p < .001$, $\eta^2 = 0.52$]. Detection accuracy for 1UF and 2UF is similarly high but decreased significantly with each additional face, and no differences were found between 6UF and 7UF. Accuracy for *Inverted Faces* followed the same pattern [F (1, 1092) = 161.77 $p < .001$, $\eta^2 = 0.51$], with no differences between 1IF and 2IF, or 6IF and 7IF. Accuracy for *Upright Non-Faces* also decreased significantly with each Set Size [F (1, 1092) = 128.17, $p < .001$, $\eta^2 = 0.45$], but no differences were found between 1UN and 2UN, and between 5UN, 6UN, and 7UN. The same significant decrease in accuracy pattern was seen for *Inverted Non-Faces* as Set Size increased - [F (1, 1092) = 127.94, $p < .001$, $\eta^2 = 0.45$]. No differences were found between 1IN and 2IN, and between 5IN, 6IN, and 7IN

Simple main effects also revealed no difference in detection accuracy at Set Size One between any condition, [F (3, 1092) = 0.06, $p = .979$, $\eta^2 = 0.00$]. A significant difference was found at Set Size Two, but further Tukey's HSD tests revealed no differences between conditions [F (3, 1092) = 3.00, p = .030, η^2 = 0.01]. At Set Sizes Three and Four, significant differences were found based on category, such that *Upright* and *Inverted Faces* were detected with similarly greater accuracy than *Upright* and *Inverted Non-Faces*, (Set Size Three [F (3, 1092) = 6.76, p < .001, η^2 = 0.02], Set Size Four [F (3, 1092) = 10.44, p < .001, η^2 = 0.03]). At Set Size Five, accuracy for *Inverted Faces* was significantly higher than both *Upright* and *Inverted Non-Faces*, but not *Upright Faces* [F (3, 1092) = 5.49, p < .001, η^2= 0.02]. At Set Size Six, significant simple main effects were found but further Tukey's HSD tests revealed no differences between conditions [F (3, 1092) = 2.36, p < .070, η^2 = 0.01]. No significant differences were found at Set Size Seven [F (3, 1092) = 1.97, p < .116, η^2 = 0.01]. At Set Size Eight, significant differences were found based on category again, however, now *Upright* and *Inverted Non-Faces* were detected with greater accuracy than *Upright* and *Inverted Faces*.

A.1.1.2 Detection Reaction Time

To investigate detection reaction time, a two-way ANOVA with repeated measures of Set Size (*One, Two*, *Three, Four, Five, Six, Seven, Eight*) and Item Type (*Upright Face, Inverted Face, Upright Non-Face, Inverted Non-Face*) was conducted. Table A.2 displays RT means and SE. The analysis revealed a significant main effect of Set Size, with RT increasing as the Set Size increased [F (7, 273) = 44.87, p < .001, η^2 = 0.53]. No main effect of Item Type was found, [F (3, 117) = 2.23, p = .089, η^2 = 0.05]. However, a significant interaction effect between Set Size and Item Type was revealed [F (21, 819) = 1.63, p = .036, η^2 = 0.04].

Table A.2 *Experiment 1 mean reaction time data in ms for Upright Face, Inverted Face, Upright Non-Face, and Inverted Non-Face conditions at each Set Size. Within-subjects standard error in brackets (Cousineau, 2005).*

Set Size	Upright Faces	Inverted Faces	Upright Non-Faces	Inverted Non-Faces	Grand Mean
One	686 (39)	686 (39)	699 (42)	706 (39)	694 (40)
Two	789 (35)	810 (37)	810 (39)	826 (34)	809 (36)
Three	941 (32)	959 (35)	1008 (35)	965 (36)	968 (35)
Four	1065 (29)	1055 (27)	1074 (36)	1058 (31)	1063 (31)
Five	1334 (41)	1287 (38)	1270 (28)	1274 (28)	1291 (34)
Six	1371 (41)	1372 (42)	1283 (38)	1310 (42)	1334 (41)
Seven	1457 (88)	1435 (56)	1334 (71)	1328 (53)	1389 (67)
Eight	1449 (82)	1329 (77)	1320 (66)	1260 (52)	1340 (69)
Grand Mean	1137 (48)	1117 (44)	1100 (44)	1091 (39)	1111 (44)

Simple main effects revealed significant increases in RTs for all conditions as Set Size increased. Reaction times for both *Upright* and *Inverted Faces* followed the same pattern [F (7, 1092) = 40.19, $p < .001$, $\eta^2 = 0.20$, and F (7, 1092) = 33.37, $p < .001$, $\eta^2 = 0.18$, respectively]. For both conditions, RTs for smaller Set Sizes between One and Four were not significantly different from their immediate neighbouring Set Size. However, RTs for these smaller Set Sizes were overall significantly quicker than RTs for larger Set Sizes between Five and Eight, which were not different from each other. The same basic pattern mentioned above was also found for *Upright* and *Inverted Non-Faces* [F (7, 1092) = 25.53, $p < .001$, $\eta^2 = 0.14$, and F (7, 1092) = 24.39, $p < .001$, $\eta^2 = 0.14$, respectively]. But in addition to this pattern, 4 *Upright Non-Faces* were not significantly different from 5 and 6 *Upright Non-Faces*. Whilst 4 *Inverted Non-Faces* were not significantly different from 8 *Inverted Non-Faces*.

Simple main effects revealed no significant differences between any Item Types at Set Sizes One to Six. However, at Set Size Seven, *Upright Faces* were significantly slower than either *Non-Face* condition. Whilst at Set Size Eight, *Upright Faces* were significantly slower than all other conditions Set Size One, F (3, 1092) = 0.09, p = .965, η^2 = 0.00; Set Size Two, F (3, 1092) = 0.21, p = .887, η^2 = 0.00; Set Size Three, F (3, 1092) = 0.74, p = .530, η^2 = 0.00; Set Size Four, F (3, 1092) = 0.07, p = .978, η^2 = 0.00; Set Size Five, F (3, 1092) = 0.81, p = .490, η^2 = 0.00; Set Size Six, F (3, 1092) = 1.87, p = .134, η2 = 0.01; Set Size Seven, F (3, 1092) = 4.21, p = .006, η^2 = 0.01; Set Size Eight, F (3, 1092) = 5.86, p = .001, η^2 = 0.02].

A.1.2 Experiment 2: Absolute Subitizing: Until Response

A.1.2.1 Detection Accuracy

To investigate detection accuracy, a two-way ANOVA with repeated measures of Set Size (*One, Two*, *Three, Four, Five, Six, Seven, Eight*) and Item Type (*Upright Face, Inverted Face, Upright Non-Face, Inverted Non-Face*) was conducted. Table A.3 displays accuracy means and SE. The analysis revealed a significant main effect of Set Size, with accuracy decreasing as the Set Size increased [F (7, 413) = 39.05, $p < .001$, $\eta^2 = 0.40$]. No main effect of Item Type was found, [F (3, 413) = 2.23, $p = .087$, $\eta^2 = 0.04$]. No significant interaction effect between Set Size and Item Type was found either [F (21, 1239) = 0.71, $p < .892$, $\eta^2 = 0.01$].

Table A.3 *Experiment 2 mean % accuracy data for Upright Face, Inverted Face, Upright Non-Face, and Inverted Non-Face conditions at each Set Size. Within-subjects standard error in brackets (Cousineau, 2005).*

Set Size	Upright Faces	Inverted Faces	Upright Non-Faces	Inverted Non-Faces	Grand Mean
One	98.82 (1.10)	98.68 (1.05)	98.54 (1.09)	98.68 (1.13)	98.68 (1.09)
Two	96.32 (0.94)	96.53 (1.03)	97.08 (1.14)	96.88 (1.18)	96.70 (1.07)
Three	95.76 (1.00)	95.00 (0.98)	95.14 (1.04)	95.21 (0.99)	95.37 (1.00)
Four	93.96 (1.05)	94.31 (1.08)	95.00 (0.96)	93.40 (0.87)	94.14 (0.99)
Five	89.38 (1.14)	88.75 (1.21)	89.45 (0.93)	88.96 (1.25)	89.14 (1.13)
Six	82.36 (1.33)	80.97 (1.29)	81.88 (1.31)	80.69 (1.51)	81.48 (1.36)
Seven	82.64 (1.41)	82.43 (1.36)	85.00 (1.59)	82.64 (1.49)	82.57 (1.46)
Eight	81.53 (1.69)	78.82 (1.91)	81.46 (1.94)	80.56 (1.91)	80.59 (1.86)
Grand Mean	90.10 (1.26)	88.64 (1.24)	90.59 (1.25)	89.09 (1.29)	89.83 (1.25)

A.1.2.2 Detection Reaction Time

To investigate detection reaction time, a two-way ANOVA with repeated measures of Set Size (*One, Two*, *Three, Four, Five, Six, Seven, Eight*) and Item Type

(*Upright Face, Inverted Face, Upright Non-Face, Inverted Non-Face*) was conducted. Table A.4 displays RT means and SE. The analysis revealed a significant main effect of Set Size, with RTs increasing as the Set Size increased [$F\ (7, 413) = 242.10$, $p < .001$, $\eta^2 = 0.80$]. A main effect of Item Type was also found, [$F\ (3, 413) = 3.04$, $p = .031$, $\eta^2 = 0.05$], with longer RTs for *Inverted Non-Faces*, but similar RTs for the remaining conditions. However, no significant interaction effect between Set Size and Item Type was found [$F\ (21, 1239) = 1.04$, $p < .416$, $\eta^2 = 0.02$].

Table A.4 *Experiment 2 mean reaction time data in ms for Upright Face, Inverted Face, Upright Non-Face, and Inverted Non-Face conditions at each Set Size. Within-subjects standard error in brackets (Cousineau, 2005).*

Set Size	Upright Faces	Inverted Faces	Upright Non-Faces	Inverted Non-Faces	Grand Mean
One	753 (39)	747 (38)	753 (38)	788 (53)	760 (42)
Two	868 (35)	877 (35)	866 (35)	896 (37)	877 (36)
Three	1030 (30)	1032 (30)	1028 (30)	1078 (32)	1042 (31)
Four	1208 (26)	1179 (27)	1207 (26)	1223 (23)	1204 (26)
Five	1673 (23)	1690 (28)	1657 (24)	1673 (22)	1673 (24)
Six	2026 (37)	2024 (39)	1999 (38)	2027 (39)	2019 (38)
Seven	2152 (46)	2169 (48)	2141 (47)	2168 (51)	2157 (48)
Eight	2068 (51)	2096 (50)	2112 (50)	2178 (56)	2113 (52)
Grand Mean	1472 (36)	1477 (37)	1470 (36)	1504 (39)	1481 (37)

A.1.3 Experiment 3: Categorical Subitizing

A.1.3.1 Detection Accuracy

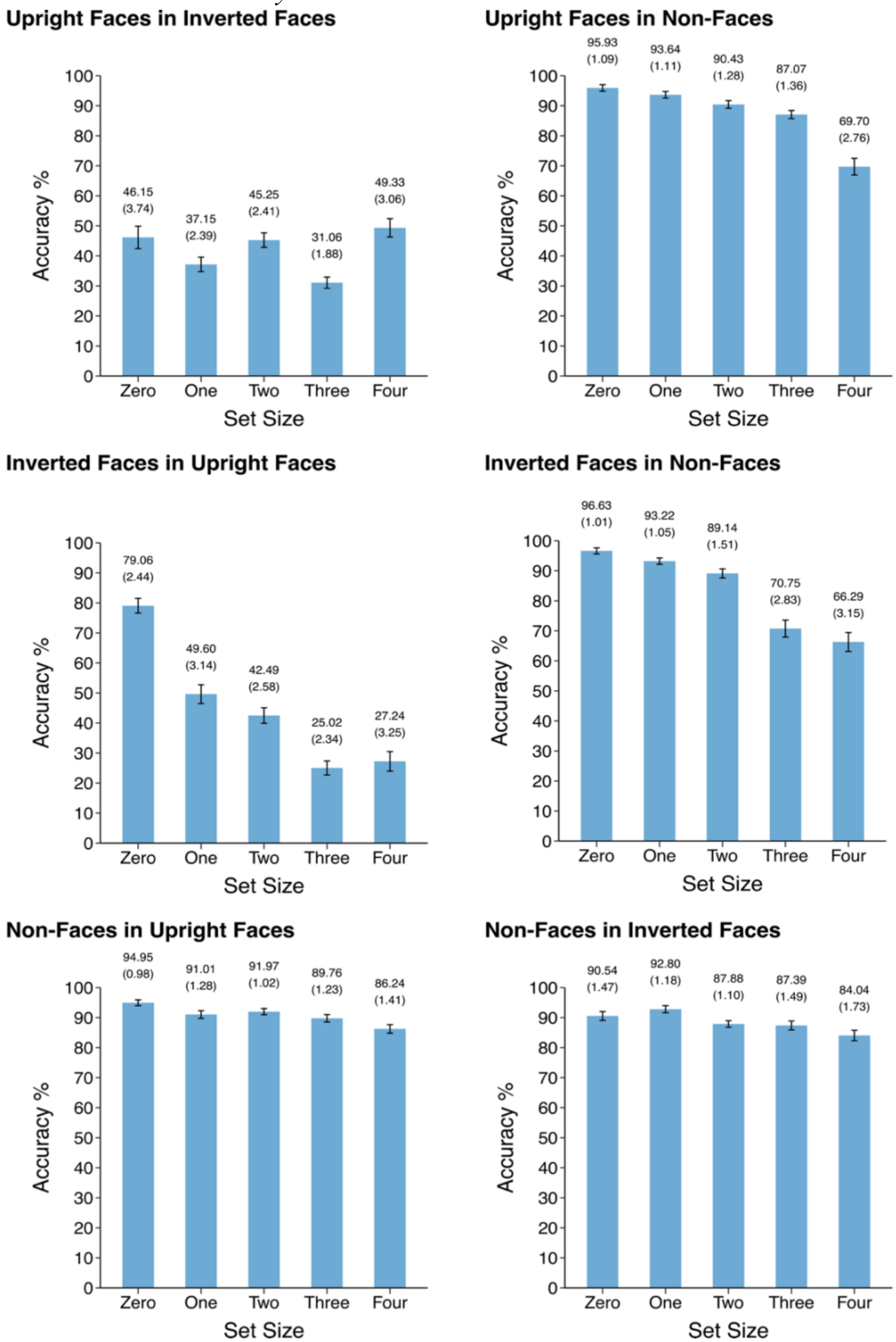

*Figure A.*1 *Mean % accuracy data for each condition in Experiment 3. Error bars show within-subjects standard error (Cousineau, 2005). Bar labels display individual condition means and (SE).*

To investigate detection accuracy, the data were subjected to a two-way ANOVA with repeated measures of Set Size (*Zero*, *One, Two*, *Three, Four*) and TARGET–distractor Type (*UPRIGHT–inverted*; *UPRIGHT–non-face*; *INVERTED–upright*; *INVERTED–non-face*; *NON-FACE–upright*; *NON-FACE–inverted*). The analysis revealed a significant main effect of Set Size whereby accuracy became poorer as Set Size increased, [$F(4, 156) = 70.48$, $p < .001$, $\eta^2 = 0.64$; Zero, M = 83.88%, SE = 1.79%; One, M = 76.24%, SE = 1.69%; Two, M = 74.53%, SE = 1.65%; Three, M = 65.17%, SE = 1.86%; Four, M = 63.80%, SE = 2.56%). A significant main effect of TARGET–distractor Type was also found [$F(4, 156) =496.73$, $p < .001$, $\eta^2 = 0.93$], with the poorest accuracy for *UPRIGHT–inverted* (M = 41.79%, SE = 2.70%) and *INVERTED–upright* (M = 44.68%, SE = 2.75%) conditions compared to other conditions (*UPRIGHT–non-face*, M = 87.35%, SE = 1.52%; *INVERTED–non-face*, M = 83.21%, SE = 1.91%; *NON-FACE–upright*, M = 90.78%, SE = 1.18%; *NON-FACE–inverted*, M = 88.53%, SE = 1.39%). A significant interaction effect between Set Size and TARGET–distractor Type was also found [$F(20, 780) = 21.85$, $p < .001$, $\eta^2 = 0.36$].

Simple main effects revealed significant differences in accuracy between the Set Sizes for all conditions except *NON-FACE–upright* [$F(4, 936) = 2.36$, $p = .052$, $\eta^2 = 0.01$]. At *UPRIGHT–inverted* [$F(4, 936) = 13.04$, $p < .001$, $\eta^2 = 0.05$], Set Sizes One and Three were significantly lower than the other Set Sizes. At *INVERTED–upright* [$F(4, 936) = 110.47$, $p < .001$, $\eta^2 = 0.32$], accuracy was highest for Set Size Zero and decreased significantly as Set Size increased, however Set Sizes One and Two were not significantly different from each other and Set Sizes Three and Four were also not significantly different from each other. At *UPRIGHT–non-face* [$F(4, 936) = 25.23$, $p < .001$, $\eta^2 = 0.1$] accuracy was the lowest at Set Size Four, but no other significant differences were found. At *INVERTED–non-face* [$F(4, 936) = 43.96$, $p < .001$, $\eta^2 = 0.16$] accuracy was similarly significantly lower for Set Sizes Three and Four. At *NON-FACE–inverted* [$F(4, 936) = 2.57$, $p = .037$, $\eta^2 = 0.01$] only accuracy at Set Size One was significantly higher than Set Size Four and no other differences were found.

Simple main effects also revealed significant differences between TARGET–distractor Types at each Set Size. At Set Size Zero [$F\ (5, 975) = 90.31$, $p < .001$, $\eta^2 = 0.32$] and Set Size One [$F\ (5, 975) = 156.10$, $p < .001$, $\eta^2 = 0.44$], accuracy for *UPRIGHT–inverted* was the lowest, and both *UPRIGHT–inverted* and *INVERTED–upright* were significantly lower than all other conditions. At Set Size Two [$F\ (5, 975) = 133.12$, $p < .001$, $\eta^2 = 0.41$] neither *UPRIGHT–inverted* and *INVERTED–upright* were different from each other but accuracy for both conditions was again lower than the remaining conditions. At Set Size Three [$F\ (5, 975) = 206.10$, $p < .001$, $\eta^2 = 0.51$] accuracy for *UPRIGHT–non-face*, *NON-FACE–upright*, and *NON-FACE–inverted* was significantly higher compared to the other conditions. At Set Size Four [$F\ (5, 975) = 117.53$, $< = .001$, $\eta^2 = 0.38$], accuracy for *UPRIGHT–non-face* and *INVERTED–non-face* was similarly lower than all other conditions, whilst accuracy at *NON-FACE–upright*, and *NON-FACE–inverted* was not different.

A.1.3.2 Detection Reaction Time

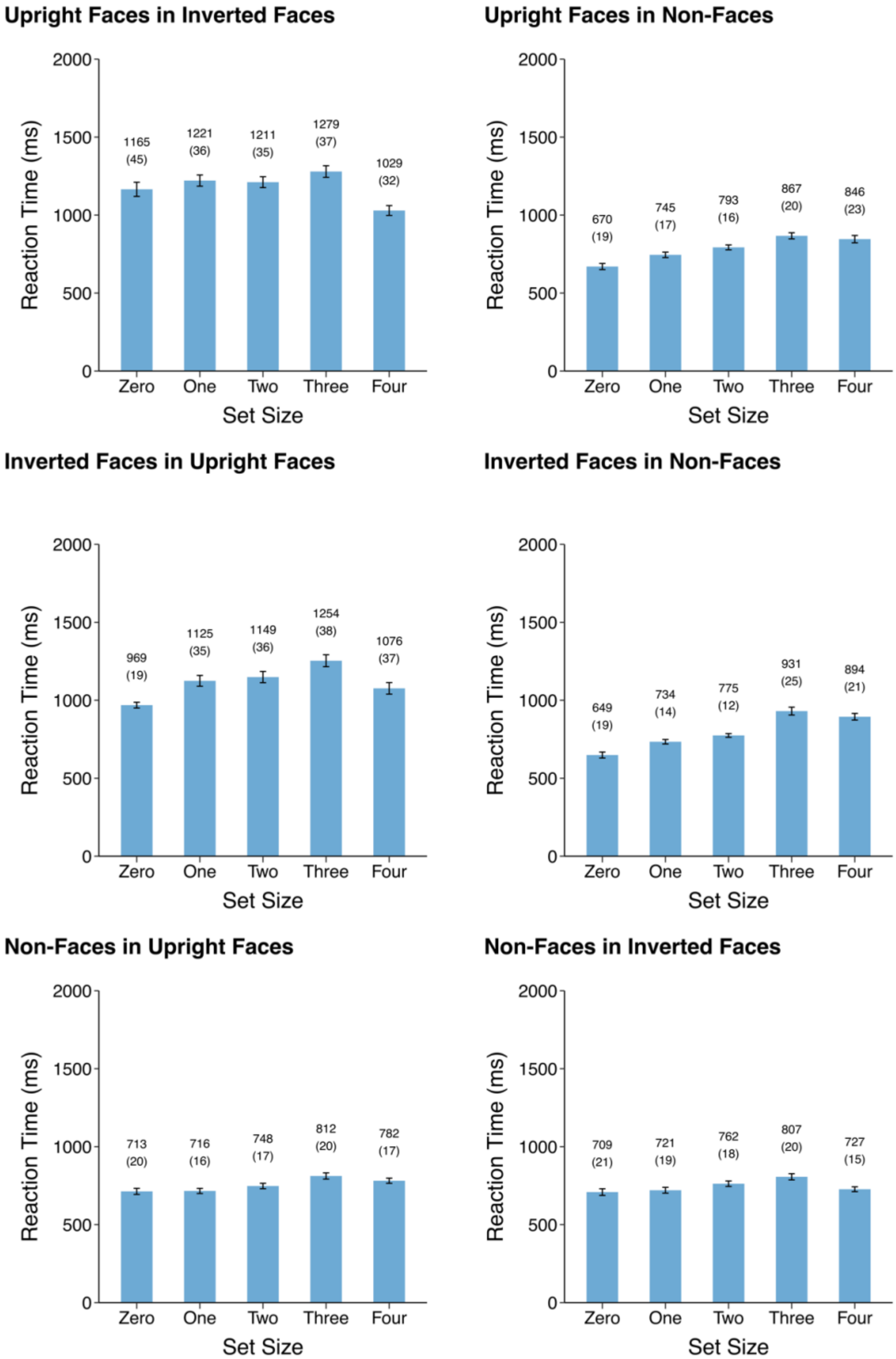

Figure A.2 Mean RT [ms] data for each condition in Experiment 3. Error bars show within-subjects standard error (Cousineau, 2005). Bar labels display individual condition means and (SE).

To investigate detection reaction, the data were subjected to a two-way ANOVA with repeated measures of Set Size (*Zero*, *One, Two*, *Three, Four*) and TARGET–distractor Type (*UPRIGHT–inverted*; *UPRIGHT–non-face*; *INVERTED–upright*; *INVERTED–non-face*; *NON-FACE–upright*; *NON-FACE–inverted*). The analysis revealed a significant main effect of Set Size whereby reaction times became slower as Set Size increased, [$F (4, 156) = 30.48$, $p < .001$, $\eta^2 = 0.44$; Zero, M = 813 ms, SE = 24 ms; One, M = 877 ms, SE = 23 ms; Two, M = 906 ms, SE = 22 ms; Three, M = 992 ms, SE = 27 ms] but then RTs became faster at Set Size Four [M = 892 ms, SE = 24 ms]. A significant main effect of TARGET–distractor Type was also found [$F (5, 195) = 119.73$, $p < .001$, $\eta^2 = 0.75$], with the slowest RTs for *UPRIGHT–inverted* (M = 1181 ms, SE = 37 ms) and *INVERTED–upright* (M = 1114 ms, SE = 33 ms) conditions compared to other conditions, (*UPRIGHT–non-face*, M = 784 ms, SE = 19 ms; *INVERTED–non-face*, M = 796 ms, SE = 18 ms; *NON-FACE–upright*, M = 754 ms, SE = 18 ms; *NON-FACE–inverted*, M = 745 ms, SE = 19 ms). A significant interaction effect between Set Size and TARGET–distractor Type was also found [$F (20, 780) = 8.66$, $p < .001$, $\eta^2 = 0.18$].

Simple main effects revealed significant differences in accuracy between the Set Sizes for all conditions except *NON-FACE–inverted* [$F (4, 936) = 3.30$, $p = .011$, $\eta^2 = 0.01$]. For *UPRIGHT–inverted* [$F (4, 936) = 18.40$, $p < .001$, $\eta^2 = 0.07$], RT at Set Size Zero was slightly slower than Set Size Three. Reaction time at Set Size Four was also significantly faster than all other Set Sizes, and no other differences were found. For the *INVERTED–upright* condition [$F (4, 936) = 22.46$, $p < .001$, $\eta^2 = 0.09$], RTs at Set Size Zero and Set Size Three were significantly different from all the other conditions, but Set Size Three RTs were the slowest. At the *UPRIGHT–non-face* [$F (4, 936) = 13.13$, $p < .001$, $\eta^2 = 0.05$] condition RTs for Set Sizes Zero and One were all similarly faster than Set Sizes Two, Three and Four, which were not significantly different from each other. A similar same pattern was seen for *INVERTED–non-face* [$F (4, 936) = 27.87$, $p < .001$, $\eta^2 = 0.11$], whereby Set Sizes Zero, One, and Two were similarly faster than Set Sizes Three and Four, which did not differ from each other. At *NON-FACE–upright* [F (4, 936) = 3.77, p = .005, η^2 = 0.02] RTs for Set Size Three were slower than Set Sizes Zero and One but no other differences were found.

Simple main effects also revealed significant differences between TARGET–distractor Types at each Set Size. At Set Size Zero [F (5, 975) = 65.95, p < .001, η^2 = 0.25], Set Size One [F (5, 975) = 81.65, p < .001, η^2 = 0.3], and Set Size Two [F (5, 975) = 69.37, p < .001, η^2 = 0.26], RTs for *UPRIGHT–inverted* and *INVERTED–upright* were similarly slower than all other conditions. The same pattern was seen in the remaining Set Sizes. However at, both Set Size Three [F (5, 975) = 72.19, p < .001, η^2 = 0.27] and Set Size Four [F (5, 975) = 28.76, p < .001, η^2 = 0.13] the INVERTED–non-face condition was also significantly slower. Across all Set Sizes, no differences between *UPRIGHT–non-face*, *NON-FACE–upright,* and *NON-FACE–inverted* were found.

A.2 Supplementary Analyses: Chapter 3

This section contains the supplementary material for Chapter 3: Capacity Limits in Face Detection, including full accuracy and reaction time analyses of Experiment 4 – 7.

A.2.1 Experiment 4: Two-vs-Three Faces and Non-Faces (Dissimilar)

A.2.1.1 Detection Accuracy

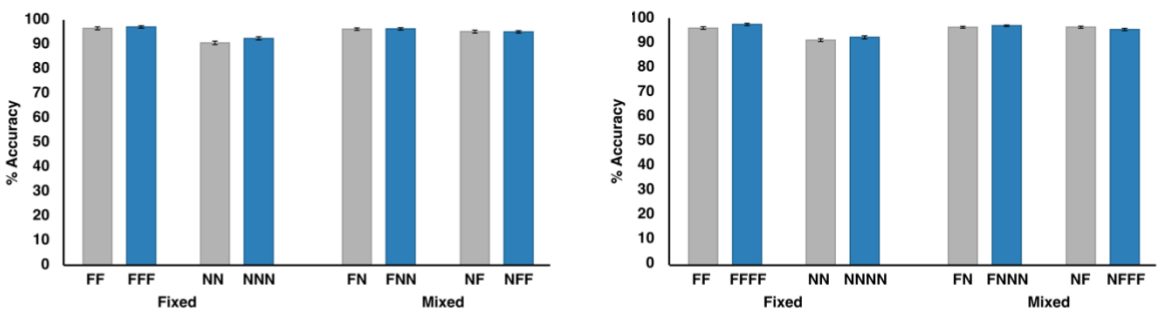

Figure A.3 Mean accuracy scores for each condition in (a) Experiment 4 and (b) Experiment 5. F denotes face, N denotes non-face. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.3a summarises the accuracy scores for each condition of Experiment 4. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Three*) and Display Type (*Fixed*, *Mixed*) revealed no main effect of Set Size with no differences in accuracy between *Two* items (M = 94.72%, SE = 0.61%) and *Three* items (M = 95.29%, SE = 0.54%) overall [F (1, 59) = 2.04, p=.158, η^2 = 0.03]. A significant main effect of Display Type was found with slightly lower detection accuracy for *Fixed* displays (M = 94.72%, SE = 0.61%) compared to *Mixed* displays (M = 95.78%, SE = 0.54%), [F (3, 177) = 25.40, p<.001, η^2 = 0.30]. However, no interaction effect between Set Size and Display Type was found, [F (3, 177) = 1.17, p=.322, η^2 = 0.02].

A.2.1.2   Detection Reaction Time

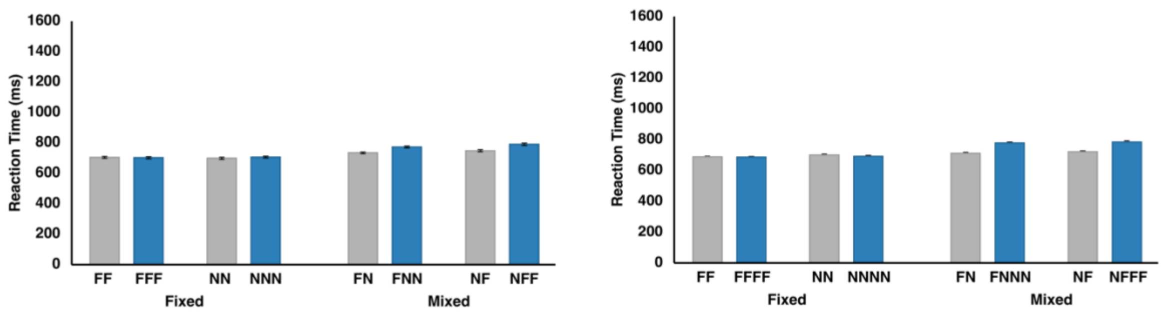

Figure A.4 Mean reaction times for each condition in (a) Experiment 4 and (b) Experiment 5. F denotes face, N denotes non-face. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.4a summarises the reaction times for each condition of Experiment 4. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Three*) and Display Type (*Fixed*, *Mixed*) revealed a significant main effect of Set Size, with faster detection times for *Two* items (M = 722ms, SE = 7ms) than for *Three* items (M = 743ms, SE = 7ms) overall, [F (1, 59) = 22.86, p<.001, η^2 = 0.28], and a significant main effect of Display Type with faster detection times for *Fixed* displays (M = 703ms, SE = 7ms) than *Mixed* displays (M = 762ms, SE = 7ms) overall, [F (3, 177) = 31.70, p<.001, η^2 = 0.35]. There was also a significant interaction between Set Size and Display Type [F (3, 177) = 6.14, p=.001, η^2 = 0.09].

Simple main effects revealed no detection time costs per additional item for *Fixed* displays when adding an upright face [FF, M = 705ms, SE = 6ms; FFF, M = 703ms, SE = 7ms; F (1, 263) = 0.05, p=.821, η^2 = 0.00], or when adding a non-face [NN, M = 699ms, SE = 7ms; NNN, M = 707ms, SE = 7ms; F (1, 263) = 0.89, p=.346, η^2 = 0.00].

However, detection time costs per additional item were found in *Mixed* displays when adding an upright face [NF, M = 749ms, SE = 8ms, NFF, = 791ms, SE = 9ms; F (1,

263) = 22.70, $p<.001$, $\eta^2 = 0.09$], and when adding a non-face [FN, M = 735ms, SE = 5ms; FNN, M = 773ms, SE = 6ms, $F(1, 263) = 18.35$, $p<.001$, $\eta^2 = 0.07$].

The effect of Display Type was significant for Set Size *Two* [$F(3, 354) = 10.09$, $p<.001$, $\eta^2 = 0.08$] and Set Size *Three* [$F(3, 354) = 35.85$, $p<.001$, $\eta^2 = 0.23$]. Separate t-tests were run to directly compare the detection of faces to non-faces within each Set Size. No differences in detection time were found for faces over non-faces in Set Size *Two,* [$t(59) = 0.55$, $p=.586$] or in Set Size *Three* [$t(59) = -0.44$, $p=.660$].

A.2.2 Experiment 5: Two-vs-Four Faces and Non-Faces (Dissimilar)

A.2.2.1 Detection Accuracy

Figure A.3b summarises the accuracy scores for each condition of Experiment 5. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Four*) and Display Type (*Fixed*, *Mixed*) revealed no main effect of Set Size with no differences in accuracy between *Two* items (M = 94.99%, SE = 0.47%) and *Four* items (M = 95.55%, SE = 0.47%) overall [F (1, 59) = 2.84, p=.097, $\eta^2 = 0.05$]. A significant main effect of Display Type was found with slightly lower detection accuracy for *Fixed* displays (M = 94.17%, SE = 0.54%), compared to *Mixed* displays (M = 96.38%, SE = 0.40%), [F (3, 177) = 35.53, p<.001, $\eta^2 = 0.37$]. There was also a significant interaction between Set Size and Display Type [F (3, 177) = 2.79, p=.042, $\eta^2 = 0.05$].

Simple main effects revealed greater accuracy when adding *Two* additional upright face to a *Fixed* display [FF, M = 95.93%, SE = 0.50%; FFFF, M = 97.37%, SE = 0.46%; F (1, 263) = 5.02, p=.026, $\eta^2 = 0.02$]. But no differences in accuracy were found when adding *Two* non-faces to a *Fixed* display [NN, M = 91.11%, SE = 0.57%; NNNN, M = 92.26%, SE = 0.65%; F (1, 263) = 3.13, p=.078, $\eta^2 = 0.01$].

No detection accuracy costs were found per additional item in *Mixed* displays when adding *Two* upright faces [NF, M = 96.52%, SE = 0.43%; NFFF, M = 95.56%, SE = 0.46%; F (1, 263) = 2.22, p=.138, $\eta^2 = 0.01$], or when adding *Two* non-faces [FN, M = 96.42%, SE = 0.36%; FNNN, M = 97.02%, SE = 0.33%; F (1, 263) = 0.86, p=.355, $\eta^2 = 0.00$]

The effect of Display Type was significant for Set Size *Two* [F (3, 354) = 24.94, p<.001, $\eta^2 = 0.17$] and Set Size *Four* [F (3, 354) = 20,06, p<.001, $\eta^2 = 0.15$]. Separate t-tests were run to directly compare the detection of faces to non-faces within each Set Size. Significantly greater detection accuracy was found for faces over non-faces in Set Size *Two*, [t (59) = 5.08, p<.001], and in Set Size *Four* [t (59) = 5.63, p<.001].

A.2.2.2 Detection Reaction Time

Figure A.4b summarises the reaction times for each condition of Experiment 5. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Four*) and Display Type (*Fixed*, *Mixed*) revealed a significant main effect of Set Size, with faster detection times for *Two* items (M = 711ms, SE = 6ms) than for *Four* items (M = 741ms, SE = 6ms) overall, [F (1, 59) = 43.29, p<.001, η^2 = 0.42], and a significant main effect of Display Type with faster detection times for *Fixed* displays (M = 696ms, SE = 6ms) than *Mixed* displays (M = 756ms, SE = 6ms) overall, [F (3, 177) = 52.89, p<.001, η^2 = 0.47]. There was also a significant interaction between Set Size and Display Type [F (3, 177) = 22.71, p<.001, η^2 = 0.28].

Simple main effects revealed no detection time costs per additional item for *Fixed* displays when adding *Two* upright faces [FF, M = 691ms, SE = 5ms; FFFF, M = 689ms, SE = 6ms; F (1, 263) = 0.05, p=.831, η^2 = 0.00], or when adding a non-face [NN, M = 705ms, SE = 6ms; NNNN, M = 697ms, SE = 6ms; F (1, 263) = 0.94, p=.332, η^2 = 0.00].

However, detection time costs per additional item were found in *Mixed* displays when adding *Two* upright faces [NF, M = 729ms, SE = 6ms, NFFF, = 793ms, SE = 7ms; F (1, 263) = 53.73, p<.001, η^2 = 0.19], and when adding *Two* non-faces [FN, M = 718ms, SE = 5ms; FNNN, M = 785ms, SE = 6ms, F (1, 263) = 58.79, p<.001, η^2 = 0.20].

The effect of Display Type was significant for Set Size *Two* [F (3, 354) = 6.25, p<.001, η^2 = 0.05] and Set Size *Four* [F (3, 354) = 73.15, p<.001, η^2 = 0.38]. Separate t-tests were run to directly compare the detection of faces to non-faces within each Set Size. No differences in detection time were found for faces over non-faces in Set Size *Two*, [t (59) = -1.62, p=.110] or in Set Size *Four* [t (59) = 0.90, p=.372].

A.2.3 Experiment 6: Two-vs-Three Faces and Non-Faces (Similar)

A.2.3.1 Detection Accuracy

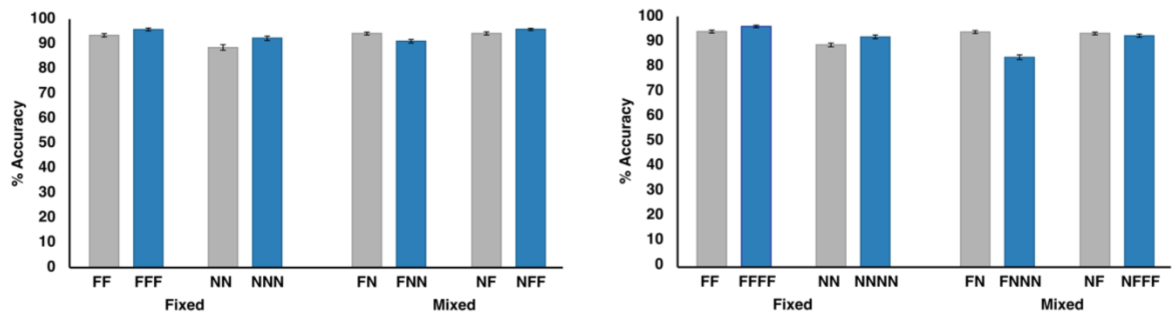

Figure A.5. Mean accuracy scores for each condition in (a) Experiment 5 and (b) Experiment 6. F denotes face, N denotes non-face. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.5a summarises the accuracy scores for each condition of Experiment 6. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Three*) and Display Type (*Fixed*, *Mixed*) revealed a significant main effect of Set Size, with slightly lower detection accuracy for *Two* items (M = 93.63%, SE = 0.73%) compared to *Three* items (M = 94.82%, SE = 0.67%) overall [F (1, 59) = 8.06, p=.006, η^2 = 0.12], and a significant main effect of Display Type, with slightly lower detection accuracy for *Fixed* displays (M = 93.54%, SE = 0.80%) compared to *Mixed* displays (M = 94.91%, SE = 0.61%) overall [F (3, 177) = 10.01, p<.001, η^2 = 0.15]. There was also a significant interaction between Set Size and Display Type [F (3, 177) = 12.97, p<.001, η^2 = 0.18].

Simple main effects revealed greater detection accuracy when adding an additional upright face to a *Fixed* display [FF, M = 94.53 %, SE = 0.66%; FFF, M = 96.91%, SE = 0.53%; F (1, 263) = 8.17, p=.005, η^2 = 0.03]. As well as greater detection accuracy when adding a non-face to a *Fixed* display [NN, M = 89.46 %, SE = 1.10%; NNN, M = 93.25%, SE = 0.90%; F (1, 263) = 20.72, p<.001, η^2 = 0.08].

However, detection accuracy was slightly higher when adding an upright face to a *Mixed* display [NF, M = 95.27%, SE = 0.59%; NFF, M = 96.95%, SE = 0.49%; F

(1, 263) = 4.06, *p*=.045, $\eta^2 = 0.02$], But, when adding a non-face to a *Mixed* display detection accuracy decreased [FN, M = 95.27%, SE = 0.59%; FNN, M = 92.15%, SE = 0.76%; F (1, 263) = 14.03, *p*<.001, $\eta^2 = 0.06$]

The effect of Display Type was significant for Set Size *Two* [F (3, 354) = 12.09, *p*<.001, $\eta^2 = 0.09$] and Set Size *Three* [F (3, 354) = 9.50, *p*<.001, $\eta^2 = 0.07$]. Separate t-tests were run to directly compare the detection of faces to non-faces within each Set Size. Significantly greater detection accuracy was found for faces over non-faces in Set Size *Two*, [*t* (59) = 3.54, *p*=.001] and in Set Size *Three* [*t* (59) = 3.01, *p*=.004].

A.2.3.2 Detection Reaction Time

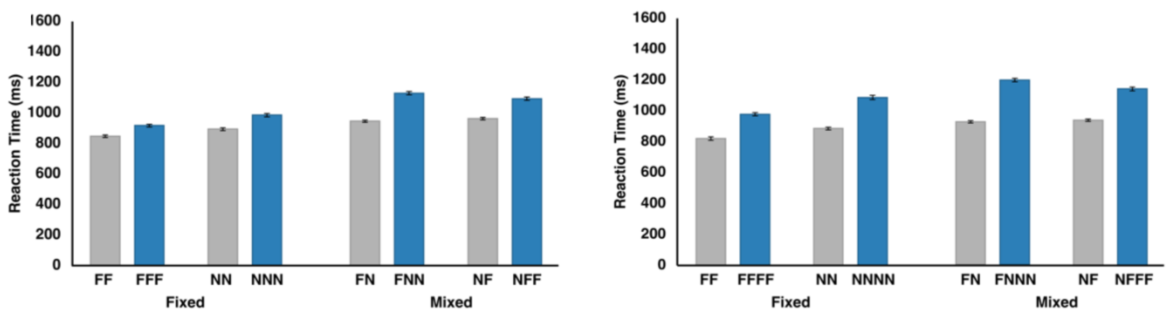

Figure A.6. Mean reaction times for each condition in (a) Experiment 6 and (b) Experiment 7. F denotes face, N denotes non-face. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.6a summarises the reaction times for each condition of Experiment 6. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Three*) and Display Type (*Fixed*, *Mixed*) revealed a significant main effect of Set Size, with faster detection times for *Two* items (M = 914ms, SE = 8ms) than for *Three* items (M = 1033ms, SE = 10ms) overall, [F (1, 59) = 240.84, *p*<.001, $\eta^2 = 0.80$], and a significant main effect of Display Type with faster detection times for *Fixed* displays (M = 912ms, SE = 9ms) than *Mixed* displays (M = 1034ms, SE = 9ms) overall, [– F

(3, 177) = 96.38, p<.001, $\eta^2 = 0.62$]. There was also a significant interaction between Set Size and Display Type [F (3, 177) = 16.87, p<.001, $\eta^2 = 0.22$].

Simple main effects revealed detection time costs per additional item for *Fixed* displays when adding an upright face [FF, M = 848ms, SE = 8ms; FFF, M = 918ms, SE = 9ms; F (1, 263) = 29.02, p<.001, $\eta^2 = 0.11$], and when adding a non-face [NN, M = 895ms, SE = 9ms; NNN, M = 987ms, SE = 11ms; F (1, 263) = 49.87, p<.001, $\eta^2 = 0.17$].

Furthermore, detection time costs per additional item were found in *Mixed* displays when adding an upright face [NF, M = 965ms, SE = 8ms, NFF, = 1095ms, SE = 11ms; F (1, 263) = 100.22, p<.001, $\eta^2 = 0.30$], and when adding a non-face [FN, M = 947ms, SE = 7ms; FNN, M = 1131ms, SE = 10ms, F (1, 263) = 199.59, p<.001, $\eta^2 = 0.46$].

The effect of Display Type was significant for Set Size *Two* [F (3, 354) = 29.40, p<.001, $\eta^2 = 0.20$] and Set Size *Three* [F (3, 354) = 101.25, p<.001, $\eta^2 = 0.46$]. Separate t-tests were run to directly compare the detection of faces to non-faces within each Set Size. Significantly lower detection times were found for faces over non-faces in Set Size *Two*, [t (59) = -4.62, p<.001] and in Set Size *Three* [t (59) = -4.71, p<.001].

A.2.4 Experiment 7: Two-vs-Four Faces and Non-Faces (Similar)

A.2.4.1 Detection Accuracy

Figure A.6b summaries the Accuracy scores for each condition of Experiment 7. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Four*) and Display Type (*Fixed*, *Mixed*) revealed a main effect of Set Size greater detection accuracy for *Two* items (M = 92.40%, SE = 0.63%) compared to *Four* items (M = 90.97%, SE = 0.72%) overall [$F\ (1, 59) = 6.91$, p=.011, $\eta^2 = 0.10$]. A significant main effect of Display Type was found with greater detection accuracy for *Fixed* displays (M = 92.58%, SE = 0.65%), compared to *Mixed* displays (M = 90.79%, SE = 0.69%), [$F\ (3, 177) = 22.89$, p<.001 $\eta^2 = 0.28$]. There was also a significant interaction between Set Size and Display Type [$F\ (3, 177) = 42.64$, p<.001, $\eta^2 = 0.42$].

Simple main effects revealed greater accuracy when adding *Two* additional upright face to a *Fixed* display [FF, M = 93.87%, SE = 0.59%; FFFF, M = 95.92%, SE = 0.50%; $F\ (1, 263) = 4.52$, p=.035, $\eta^2 = 0.02$]. As well as when adding *Two* non-faces to a *Fixed* display [NN, M = 88.66%, SE = 0.76%; NNNN, M = 91.87%, SE = 0.77%; $F\ (1, 263) = 11.06$, p=.001, $\eta^2 = 0.05$].

No detection accuracy costs were found per additional item in *Mixed* displays when adding *Two* upright faces [NF, M = 93.19%, SE = 0.45%; NFFF, M = 92.30%, SE = 0.67%; $F\ (1, 263) = 0.85$, p=.356, $\eta^2 = 0.00$]. However, detection accuracy costs were found when adding *Two* non-faces [FN, M = 93.87%, SE = 0.62%; FNNN, M = 83.80%, SE = 0.93%; $F\ (1, 263) = 108.95$, p<.001 $\eta^2 = 0.32$].

The effect of Display Type was significant for Set Size *Two* [$F\ (3, 354) = 11.99$, p<.001, $\eta^2 = 0.09$] and Set Size *Four* [$F\ (3, 354) = 49.73$, p<.001, $\eta^2 = 0.30$]. Separate t-tests were run to directly compare the detection of faces to non-faces within each Set Size. Significantly greater detection accuracy was found for faces over non-faces in Set Size *Two*, [$t\ (59) = 5.43$, p<.001] and in Set Size *Four* [$t\ (59) = 4.23$, p<.001].

A.2.4.2 Detection Reaction Time

Figure A.6b summarises the reaction times for each condition of Experiment 7. A two-way ANOVA with the repeated measures factors of Set Size (*Two*, *Four*) and Display Type (*Fixed*, *Mixed*) revealed a significant main effect of Set Size, with faster detection times for *Two* items (M = 893ms, SE = 9ms) compared to *Four* items (M = 1100ms, SE = 12ms) overall, [$F(1, 59) = 333.18$, $p<.001$, $\eta^2 = 0.53$], and a significant main effect of Display Type with faster detection times for *Fixed* displays (M = 942ms, SE = 11ms) compared to *Mixed* displays (M = 1050ms, SE = 10ms) overall, [$F(3, 177) = 67.55$, $p<.001$ $\eta^2 = 0.85$]. There was also a significant interaction between Set Size and Display Type [$F(3, 177) = 18.18$, $p<.001$, $\eta^2 = 0.24$].

Simple main effects revealed detection time costs per additional item for *Fixed* displays when adding *Two* upright faces [FF, M = 820ms, SE = 11ms; FFFF, M = 977ms, SE = 12ms; $F(1, 263) = 111.68$, $p<.001$, $\eta^2 = 0.32$], and when adding a non-face [NN, M = 886ms, SE = 9ms; NNNN, M = 1087ms, SE = 14ms; $F(1, 263) = 183.41$, $p<.001$, $\eta^2 = 0.44$].

Furthermore, detection time costs per additional item were found in *Mixed* displays when adding *Two* upright faces [NF, M = 941ms, SE = 8ms, NFFF, = 1144ms, SE = 12ms; $F(1, 263) = 186.36$, $p<.001$, $\eta^2 = 0.44$], and when adding *Two* non-faces [FN, M = 923ms, SE = 8ms; FNNN, M = 1193ms, SE = 12ms, $F(1, 263) = 332.16$, $p<.001$ $\eta^2 = 0.58$].

The effect of Display Type was significant for Set Size *Two* [$F(3, 354) = 26.74$, $p<.001$, $\eta^2 = 0.18$] and Set Size *Four* [$F(3, 354) = 80.62$, $p<.001$, $\eta^2 = 0.41$]. Separate t-tests were run to directly compare the detection of faces to non-faces within each Set Size. Significantly lower detection times were found for faces over non-faces in Set Size *Two*, [$t(59) = -4.87$, $p<.001$] or in Set Size *Four* [$t(59) = -6.11$, $p<.001$].

A.3 Supplementary Analyses: Chapter 4

This section contains the supplementary material for Chapter 4: Multiple Face Detection, including full accuracy and reaction time analyses of Experiment 8 – 13.

A.3.1 Experiment 8: Upright, Inverted & Scrambled in Real Scenes (250 ms)

A.3.1.1 Detection Accuracy

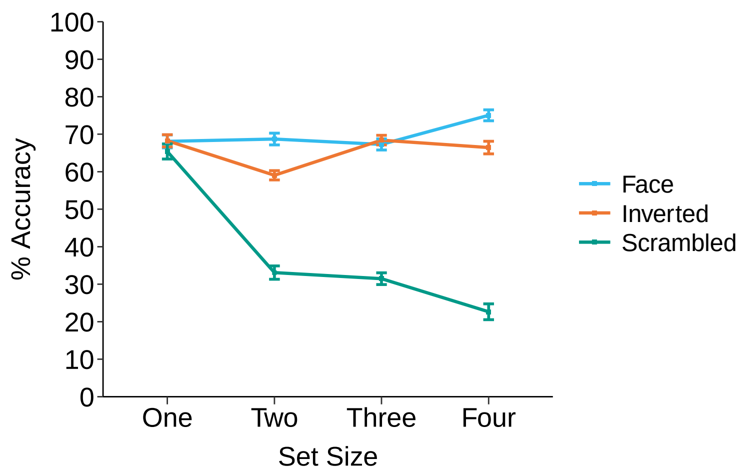

Figure A.7 Mean percentage accuracy detection for Upright, Inverted, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.7 reports accuracy data for Experiment 8 for each Target Type at each Set Size. To investigate detection accuracy, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [$F(3, 177) = 21.62$, $p < .001$, $\eta^2 = 0.27$], with highest detection accuracy for *One* Target (M = 67.24%, SE = 1.8%), but similar detection accuracy for the remaining Set Sizes, (*Two*, M = 53.62 %, SE = 1.53%; *Three*, M = 55.72 %, SE = 1.45%; *Four*, M = 54.71 %, SE = 1.75 %). There was also a significant main effect of Target Type [$F(2, 118) = 444.12$, $p < .001$, $\eta^2 = 0.88$], with higher accuracy for

Upright Targets ($M = 69.78\%$, $SE = 1.56\%$) and *Inverted* Targets ($M = 65.54\%$, $SE = 1.46\%$) compared to *Scrambled* Targets ($M = 38.15\%$, $SE = 1.87\%$). The interaction effect between Set Size and Target Type was also significant [$F\ (6, 354) = 73.40$, $p < .001$, $\eta^2 = 0.55$].

Simple main effects revealed significant differences in accuracy across Set Size for each Target Type. At the *Upright* condition [$F\ (3, 531) = 4.15$, $p = .006$, $\eta^2 = 0.02$], accuracy for Four Upright Targets ($M = 75.04\ \%$, $SE = 1.46\ \%$) was significantly higher than One Upright ($M = 68.09\%$, $SE = 1.74\%$) and Three Upright ($M = 67.28\ \%$, $SE = 1.49\%$) but not Two Upright ($M = 68.72\ \%$, $SE = 1.57\%$). At the *Inverted* condition [$F\ (3, 1.57) = 6.40$, $p < .001$, $\eta^2 = 0.03$] accuracy for Two Inverted Targets was significantly lower than all other Set Sizes [One Inverted, $M = 68.24\%$, $SE = 1.64\%$; Two Inverted, $M = 59.05\ \%$, $SE = 1.24\%$; Three Inverted, $M = 68.42\ \%$, $SE = 1.29\%$; Four Inverted, $M = 66.44\ \%$, $SE = 1.68\ \%$). At the Scrambled condition [$F\ (3, 1.57) = 115.06$, $p = .001$, $\eta^2 = 0.39$] accuracy was significantly higher at One Scrambled Targets, and significantly lower at Four Scrambled Targets, (One Scrambled, $M = 65.39\%$, $SE = 2.02\%$; Two Scrambled, $M = 33.09\ \%$, $SE = 1.78\%$; Three Scrambled $M = 31.47\ \%$, $SE = 1.58\%$; Four Scrambled, $M = 22.64\ \%$, $SE = 2.11\ \%$).

There was no significant difference in accuracy between Target Types at Set Size One [$F\ (2, 472) = 71.63$, $p < .001$, $\eta^2 = 0.23$]. However, *Upright* and *Inverted* Targets were detected more efficiently than *Scrambled* Targets at every other Set Size, [Set Size Two, $F\ (2, 472) = 43.16$, $p < .001$, $\eta^2 = 0.15$; Set Size Three, $F\ (2, 472) = 26.73$, $p < .001$, $\eta^2 = 0.10$; Set Size Four, $F\ (2, 472) = 18.68$, $p < .001$, $\eta^2 = 0.07$].

A.3.1.2 Detection Reaction Time

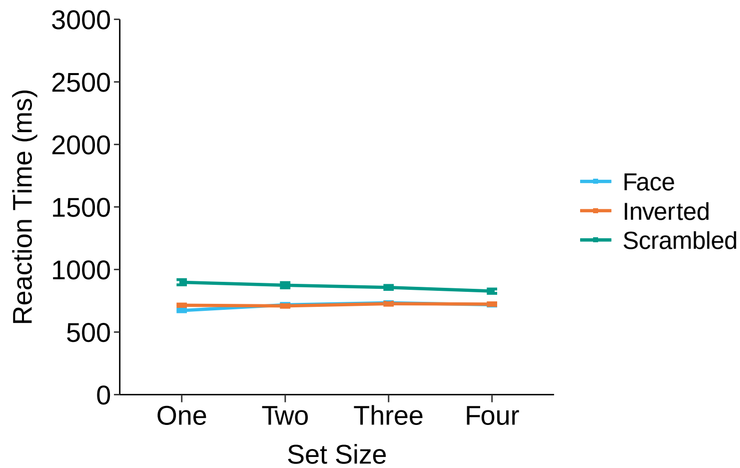

Figure A.8 Mean detection RT detection for Upright, Inverted, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.8 reports RT data for Experiment 8 for each Target Type at each Set Size. To investigate detection RT, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed no significant main effect of Set Size [F (3, 177) = 0.60, p = .618, η^2 = 0.01], with similar RTs for *One* (M = 761 ms, SE = 14 ms), *Two* (M = 767 ms, SE = 15 ms), *Three* (M = 772 ms, SE = 12 ms), and *Four* (M = 756 ms, SE = 13 ms). However, there was a significant main effect of Target Type [F (2, 118) = 97.71, p < .001, η^2 = 0.62], with faster RTs for *Upright* Targets (M = 710 ms, SE = 10 ms) and *Inverted* Targets (M = 718 ms, SE = 11 ms), compared to *Scrambled* Targets (M = 864 ms, SE = 19 ms). The interaction effect between Set Size and Target Type was also significant [F (6, 354) = 4.57, p < .001, η^2 = 0.07].

Simple main effects revealed significant differences in RT across Set Size for Target Type at the *Upright* [F (3, 531) = 3.66, p = .012, η^2 = 0.02] and *Scrambled* [F (3, 1.57) = 4.76, p = .003, η^2 = 0.03] conditions. One Upright Targets were detected significantly faster than Three Upright Targets, but no other differences were found (One Upright, M = 672 ms, SE = 10 ms; Two Upright, M = 717 ms, SE = 12 ms; Three Upright, M = 733 ms, SE = 9 ms; Four Upright, M = 718 ms, SE = 10 ms). At the *Scrambled* condition, One Scrambled Target was detected significantly slower than Four Scrambled Targets, but no other differences were found [One Scrambled, M = 898 ms, SE = 21 ms; Two Scrambled, M = 874 ms, SE = 21 ms; Three Scrambled, M = 857 ms, SE = 16 ms; Four Scrambled, M = 827 ms, SE = 17 ms). No significant differences in RT were found for the *Inverted* condition [$F\ (3, 1.57) = 0.35$, $p = .793$, $\eta^2 = 0.00$, One Inverted, M = 714 ms, SE = 10 ms; Two Inverted, M = 708 ms, SE = 11 ms; Three Inverted, M = 726 ms, SE = 12 ms; Four Inverted, M = 723 ms, SE = 11 ms].

Simple main effects also revealed significantly slower detection RTs for the *Scrambled* condition compared to *Upright* and *Inverted* at each Set Size. [Set Size One, $F\ (2, 472) = 71.63$, $p < .001$, $\eta^2 = 0.23$; Set Size Two, $F\ (2, 472) = 43.16$, $p < .001$, $\eta^2 = 0.15$; Set Size Three, $F\ (2, 472) = 26.73$, $p < .001$, $\eta^2 = 0.10$; Set Size Four, $F\ (2, 472) = 18.68$, $p < .001$, $\eta^2 = 0.07$].

A.3.2 Experiment 9: Upright, Inverted & Scrambled in Real Scenes (unlimited exposure)

A.3.2.1 Detection Accuracy

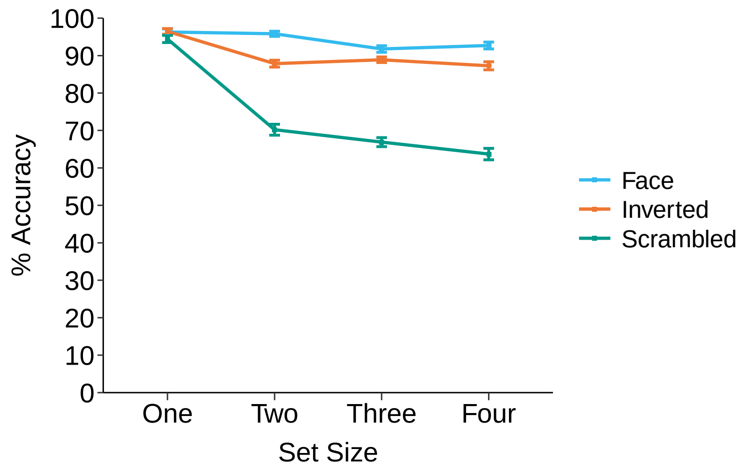

Figure A.9 Mean percentage accuracy detection for Upright, Inverted, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.9 reports accuracy data for Experiment 9 for each Target Type at each Set Size. To investigate detection accuracy, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 85.28, $p < .001$, $\eta^2 = 0.59$], with highest detection accuracy for *One* Target (M = 95.73%, SE = 0.85%), but similar detection accuracy for the remaining Set Sizes, (*Two* M = 84.62%, SE = 1.03%*; Three*, M = 82.51%, SE = 0.94%; *Four*, M = 81.22%, SE = 1.18%). There was also a significant main effect of Target Type [F (2, 118) = 231.41, $p < .001$, $\eta^2 = 0.80$], with higher accuracy for *Upright* Targets (M = 94.14%, SE = 0.83%) and *Inverted* Targets (M = 90.12%, SE = 0.88%) compared to *Scrambled* Targets (M = 73.81%, SE = 1.29%). The interaction effect between Set Size and Target Type was also significant [F (6, 354) = 63.38, $p < .001$, $\eta^2 = 0.52$].

Simple main effects revealed significant differences in accuracy across Set Size for each Target Type. At the *Upright* condition [F (3, 531) = 5.22, $p < .001$, $\eta^2 = 0.03$], accuracy for Three Upright Targets was significantly lower than One or Two Upright Targets, but no other differences were found (One Upright, M = 96.30%, SE = 0.83%; Two Upright, M = 95.83%, SE = 0.69%; Three Upright, M = 91.76%, SE = 0.87%; Four Upright, M = 92.69%, SE = 0.92%). At the *Inverted* condition [F (3, 0.69) = 18.70, $p < .001$, $\eta^2 = 0.10$] accuracy for One Inverted Targets was significantly higher than all other Set Sizes (One Inverted, M = 96.45%, SE = 0.76%; Two Inverted, M = 87.85%, SE = 0.93%; Three Inverted, M = 88.89%, SE = 0.74%; Four Inverted, M = 87.28%, SE = 1.08%). At the *Scrambled* condition [F (3, 0.69) = 200.79, $p < .001$, $\eta^2 = 0.53$] accuracy was highest for One Scrambled Target, followed by Two Scrambled Targets but no other differences were found (One Scrambled , M = 94.44%, SE = 0.96%; Two Scrambled , M = 70.19%, SE = 1.46%; Three Scrambled , M = 66.89%, SE = 1.21%; Four Scrambled , M = 63.70%, SE = 1.53%)

No significant differences in accuracy were found between the Target Types at Set Size One [F (2, 472) = 1.22, $p = .296$, $\eta^2 = 0.01$]. But at the remaining Set Sizes, accuracy for *Scrambled* Targets was significantly poorer than *Upright* or *Inverted* Targets Set Size Two [F (2, 472) = 168.46, $p < .001$, $\eta^2 = 0.42$; Set Size Three [F (2, 472) = 181.02, $p < .001$, $\eta^2 = 0.43$; Set Size Four [F (2, 472) = 232.27, $p < .001$, $\eta^2 = 0.50$].

A.3.2.2 Detection Reaction Time

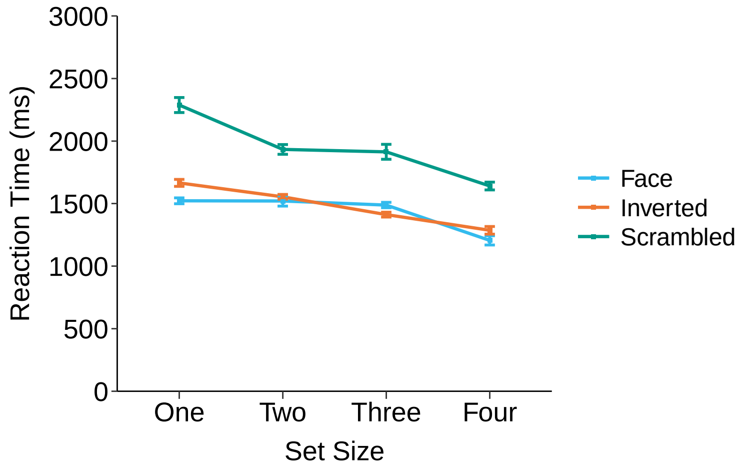

Figure A.10 Mean detection RT detection for Upright, Inverted, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.10 reports RT data for Experiment 9 for each Target Type at each Set Size. To investigate detection RT, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 48.25, $p < .001$, $\eta^2 = 0.45$], with higher RTs for *One* Target (M = 1825 ms, SE = 37 ms) compared to *Two* (M = 1669 ms, SE = 33 ms), *Three* Targets (M = 1605 ms, SE = 34 ms), while *Four* Targets were detected the fastest (M = 1378 ms, SE = 33 ms). However, there was a significant main effect of Target Type [F (2, 118) = 139.37, $p < .001$, $\eta^2 = 0.70$], with faster RTs for *Upright* Targets (M = 1434 ms, SE = 31 ms) and *Inverted* Targets (M = 1479 ms, SE = 24 ms), compared to *Scrambled* Targets (M = 1944 ms, SE = 47 ms). The interaction effect between Set Size and Target Type was also significant [F (6, 354) = 10.30, $p < .001$, $\eta^2 = 0.15$].

Simple main effects revealed significant differences in RT across Set Size for each Target Type. For the *Upright* condition [F (3, 531) = 18.2, $p < .001$, $\eta^2 = 0.09$], RTs for Four Upright Targets were significantly faster than the other Set Sizes (One Upright, M = 1522 ms, SE = 23 ms; Two Upright, M = 1521 ms, SE = 41 ms; Three

Upright, M = 1488 ms, SE = 22 ms; Four Upright, M = 1206 ms, SE = 37 ms). For the *Inverted* condition [F (3, 531) = 21.22, $p < .001$, $\eta^2 = 0.11$], RTs for One and Two Inverted Targets were similarly significantly slower than Three and Four Inverted Targets, which also did not differ from each other (One Inverted, M = 1665 ms, SE = 28 ms; Two Inverted, M = 1554 ms, SE = 19 ms; Three Inverted, M = 1412 ms, SE = 19 ms; Four Inverted, M = 1286 ms, SE = 30 ms). Whilst RTs at the *Scrambled* condition [F (3, 531) = 54.68, p = .001, η2 = 0.24] were slowest for One Scrambled Target (M = 2288 ms, SE = 60 ms), intermediate and not significantly different at Two Scrambled (M = 1933 ms, SE = 39 ms) and Three Scrambled Targets (M = 1914 ms, SE = 60 ms), and significantly faster at Four Scrambled Targets (M = 1641 ms, SE = 31 ms).

Simple main effects also revealed significantly slower detection RTs for the *Scrambled* condition compared to *Upright* and *Inverted* at each Set Size [Set Size One, F (2, 472) = 136.29, $p < .001$, $\eta^2 = 0.37$; Set Size Two, F (2, 472) = 43.11, $p < .001$, $\eta^2 = 0.15$; Set Size Three, F (2, 472) = 60.24, $p < .001$, $\eta^2 = 0.20$; Set Size Four, F (2, 472) = 43.99, $p = .001$, $\eta^2 = 0.16$].

A.3.3 Experiment 10: Upright, Inverted & Scrambled in Blank Scenes (250 ms)

A.3.3.1 Detection Accuracy

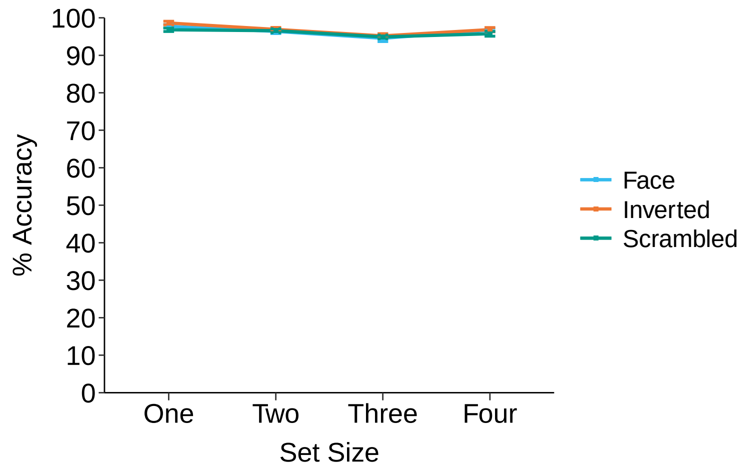

Figure A.11 Mean percentage accuracy detection for Upright, Inverted, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.11 reports accuracy data for Experiment 10 for each Target Type at each Set Size. To investigate detection accuracy, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [$F(3, 177) = 10.34$, $p < .001$, $\eta^2 = 0.15$], with lowest accuracy seen for *Three* Targets (M = 94.87%, SE = 0.65%) compared to *One* (M = 97.72%, SE = 0.46%), *Two* (M = 96.63%, SE = 0.53%), and *Four* Targets (M = 96.45%, SE = 0.55%). There was no significant main effect of Target Type [$F(2, 118) = 2.53$, $p = .084$, $\eta^2 = 0.04$], and accuracy was similarly high for all Target Types (*Upright* Targets*,* M = 96.35%, SE = 0.59%); *Inverted* Targets, M = 96.89%, SE = 0.53%; *Scrambled* Targets, M = 96.00%, SE = 0.53%) . The interaction effect between Set Size and Target Type was also non-significant [$F(6, 354) = 0.60$, $p = .734$, $\eta^2 = 0.01$].

Accuracy was at ceiling across all Set Sizes and Target Types (One Upright, M = 97.74%, SE = 0.45%; Two Upright, M = 96.38%, SE = 0.57%; Three Upright, M

= 94.53%, SE = 0.86%; Four Upright, M = 96.76%, SE = 0.46%; One Inverted, M = 98.61%, SE = 0.48%; Two Inverted, M = 96.93%, SE = 0.49%; Three Inverted, M = 95.17%, SE = 0.6%; Four Inverted, M = 96.85%, SE = 0.56%; One Scrambled , M = 96.81%, SE = 0.45%; Two Scrambled , M = 96.57%, SE = 0.52%; Three Scrambled , M = 94.90%, SE = 0.5%; Four Scrambled , M = 95.74%, SE = 0.64%).

A.3.3.2 Detection Reaction Time

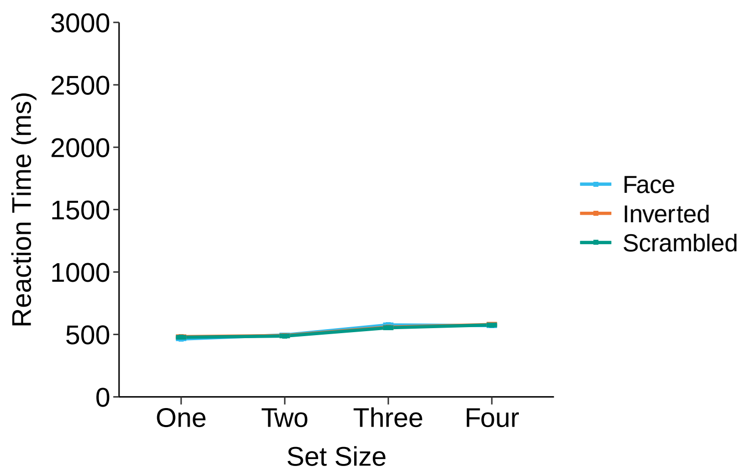

Figure A.12 Mean detection RT for Upright, Inverted, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.12 reports RT data for Experiment 10 for each Target Type at each Set Size. To investigate detection RT, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 49.52, $p < .001$, $\eta^2 = 0.46$], with faster RTs for *One* Target (M = 474 ms, SE = 7 ms) and *Two* Targets (M = 491 ms, SE = 7 ms) compared to *Three* Targets (M = 563 ms, SE = 7 ms) and *Four* Targets (M = 575 ms, SE = 9 ms). There was also a significant main effect of Target Type [F (2, 118) = 0.64, $p = .531$, $\eta^2 = 0.01$], with faster RTs for *Upright* Targets (M = 1434, SE = 31) and *Inverted* Targets (M = 1479, SE = 24), compared to *Scrambled* Targets (M = 1944, SE = 47). The interaction effect between Set Size and Target Type was also significant [F (6, 354) = 3.19, p = .005, $\eta^2 = 0.05$].

Simple main effects revealed significant differences in RT across Set Size for each Target Type. For the *Upright* condition [F (3, 531) = 45.02, $p < .001$, $\eta^2 = 0.02$], RTs for Four Upright Targets were significantly faster than One and Two Targets, but not Three Targets (One Upright, M = 464 ms, SE = 6 ms; Two Upright, M = 496 ms, SE = 7 ms; Three Upright, M = 576 ms, SE = 8 ms; Four Upright, M = 571 ms, SE = 8 ms). For the *Inverted* [F (3, 531) = 34.03, $p < .001$, $\eta^2 = 0.16$] and *Scrambled* conditions [F (3, 531) = 34.42, p = .001, η2 = 0.16], RTs for One and Two Targets were similarly significantly faster than Three and Four Targets, which also did not differ from each other (One Inverted, M = 482 ms, SE = 7 ms; Two Inverted, M = 491 ms, SE = 7 ms; Three Inverted, M = 558 ms, SE = 6 ms; Four Inverted, M = 579 ms, SE = 10 ms; One Scrambled , M = 477 ms, SE = 7 ms; Two Scrambled , M = 487 ms, SE = 7 ms; Three Scrambled , M = 554 ms, SE = 8 ms; Four Scrambled , M = 576 ms, SE = 8 ms).

A significant difference between Target Types was found at Set Size One, where *Upright* Targets were detected significantly faster than *Inverted* but not Scrambled Targets [F (2, 472) = 3.29, p = .038, $\eta^2 = 0.01$]. At Set Size Three, *Upright* Targets were detected significantly slower than *Inverted* and Scrambled Targets [F (2, 472) = 5.29, p = .005, $\eta^2 = 0.02$]. No other differences were found at either Set Size Two [F (2, 472) = 0.68, p = .507, $\eta^2 = 0.00$] or Set Size Four [F (2, 472) = 0.70, p = .495, $\eta^2 = 0.00$].

A.3.4 Experiment 11: Upright, Inverted, Scrambled in Voronoi Scenes (250 ms)

A.3.4.1 Detection Accuracy

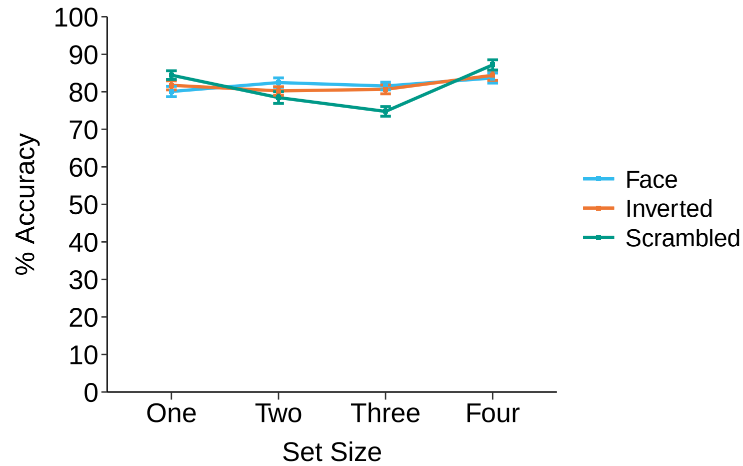

Figure A.13 Mean percentage accuracy detection for Upright, Inverted, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.13 reports accuracy data for Experiment 11 for each Target Type at each Set Size. To investigate detection accuracy, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [$F\ (3, 177) = 7.05$, $p < .001$, $\eta^2 = 0.11$], with intermediate accuracy for *One* (M = 82.07%, SE = 1.24%) and *Two* (M = 80.38%, SE = 1.32%) Targets, lowest accuracy for *Three* (M = 78.99%, SE = 1.16%) Targets, and the highest accuracy for *Four* (M = 85.09%, SE = 1.35%) Targets. There was no significant main effect of Target Type [$F\ (2, 118) = 0.37$, $p = .689$, $\eta^2 = 0.01$], and accuracy was similarly high for all Target Types (*Upright* Targets*,* M = 81.93%, SE = 1.25%; *Inverted* Targets, M = M = 81.75%, SE = 1.22%; *Scrambled* Targets, M = 81.21%, SE = 1.34%) . The interaction effect between Set Size and Target Type was also non-significant [$F\ (6, 354) = 6.57$, $p < .001$, $\eta^2 = 0.10$].

Simple main effects revealed significant differences in accuracy across Set Size only for the *Scrambled* condition [F (3, 1.26) = 17.29, p = .001, η^2 = 0.09]. Accuracy for Two and Three Scrambled Targets was similarly significantly lower than for One and Four Scrambled Targets (One Scrambled , M = 84.43%, SE = 1.16%; Two Scrambled , M = 78.46%, SE = 1.57%; Three Scrambled , M = 74.78%, SE = 1.28%; Four Scrambled , M = 87.18%, SE = 1.35%). No accuracy differences across Set Size were found for *Upright* [F (3, 531) = 1.26, p = .286, η^2 = 0.01] or *Inverted* conditions [F (3, 1.26) = 1.94, p = .123, η^2 = 0.01], (One Upright, M = 80.07%, SE = 1.36%; Two Upright, M = 82.46%, SE = 1.26%; Three Upright, M = 81.54%, SE = 1.02%; Four Upright, M = 83.67%, SE = 1.34%; One Inverted, M = 81.72%, SE = 1.20%; Two Inverted, M = 80.23%, SE = 1.12%; Three Inverted, M = 80.64%, SE = 1.18%; Four Inverted, M = 84.41%, SE = 1.37%).

Simple main effects between Target Types at each Set Size revealed significant differences at Set Sizes One and Three. At Set Size One [F (2, 472) = 3.63, p = .027, η^2 = 0.02], accuracy for One Upright Target was significantly lower than One Scrambled Target. At Set Size Three [F (2, 472) = 10.1, p < .001, η^2 = 0.04], accuracy for Three Scrambled Targets was significantly lower than Three Upright and Inverted Targets. No differences between Target Types were found at Set Size Two [F (2, 472) = 3.00, p = .051, η^2 = 0.01], or Set Size Four [F (2, 472) = 2.56, p = .079, η^2 = 0.01].

A.3.4.2 Detection Reaction Time

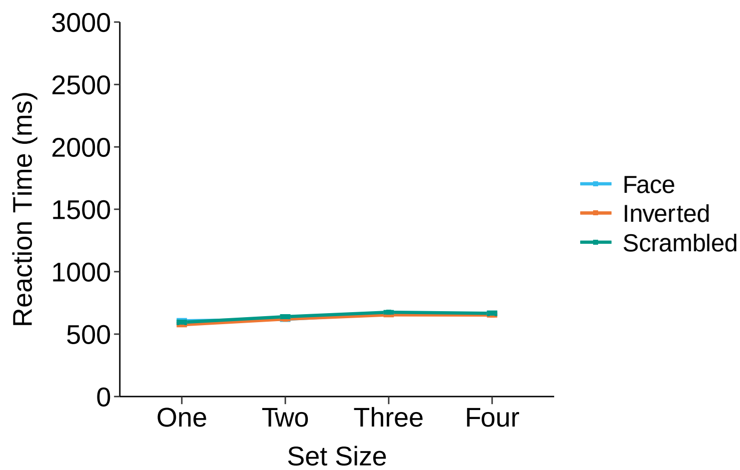

Figure A.14 Mean detection RT for Upright, Inverted, and Scrambled conditions at each Set Size.Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.14 reports RT data for Experiment 11 for each Target Type at each Set Size. To investigate detection RT, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright, Inverted, Scrambled*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 20.63, $p < .001$, $\eta^2 = 0.26$], with fastest RTs for *One* Target (M = 592 ms, SE = 10 ms) followed by *Two* Targets (M = 626 ms, SE = 8 ms) then similarly slower RTs for *Three* Targets (M = 664 ms, SE = 7 ms) and *Four* Targets (M = 662 ms, SE = 10 ms). There was also a significant main effect of Target Type [F (2, 118) = 5.99, $p = .003$, $\eta^2 = 0.09$], with faster RTs for *Inverted* Targets (M = 625 ms, SE = 8 ms) compared to *Upright* Targets (M = 639 ms, SE = 10 ms), and *Scrambled* Targets (M = 644 ms, SE = 9 ms). But the interaction effect between Set Size and Target Type was non-significant [F (6, 354) = 1.04, $p = .400$, $\eta^2 = 0.05$].

Reaction times for all Target Types across all Set Sizes were similar (One Upright, M = 606 ms, SE = 11 ms; Two Upright, M = 619 ms, SE = 10 ms; Three Upright, M = 663 ms, SE = 7 ms; Four Upright, M = 667 ms, SE = 10 ms; One Inverted, M = 575 ms, SE = 9 ms; Two Inverted, M = 620 ms, SE = 7 ms; Three

Inverted, M = 654 ms, SE = 8 ms; Four Inverted, M = 653 ms, SE = 10 ms; One Scrambled, M = 594 ms, SE = 11 ms; Two Scrambled, M = 639 ms, SE = 9 ms; Three Scrambled, M = 675 ms, SE = 7 ms; Four Scrambled, M = 666 ms, SE = 11 ms).

A.3.5 Experiment 12: Scrambled and Sideways in Real Scenes (250 ms)

A.3.5.1 Detection Accuracy

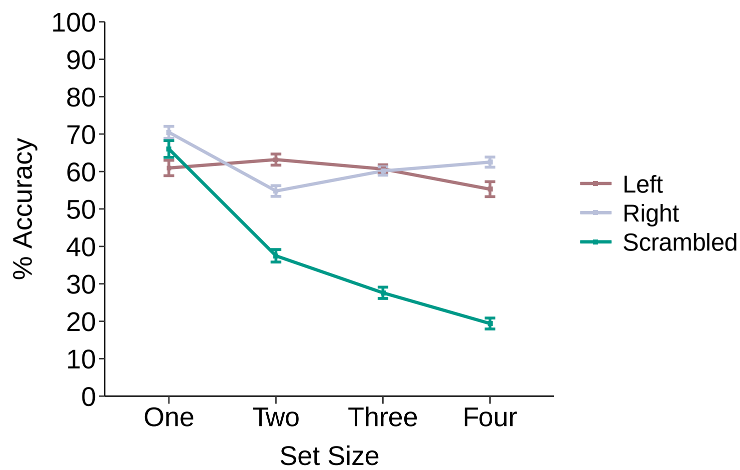

Figure A.15 Mean percentage detection accuracy for Left, Right, and Scrambled conditions at each Set Size.Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.15 reports accuracy data for Experiment 12 for each Target Type at each Set Size. To investigate detection accuracy, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Left, Right, Scrambled*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 38.78, $p < .001$, $\eta^2 = 0.40$], accuracy decreasing as Set Size increased, (*One*, M = 65.81%, SE = 1.97%; *Two*, M = 51.83%, SE = 1.53%; *Three*, M = 49.47%, SE = 1.27%; *Four*, M = 45.73%, SE = 1.6%). There was also a significant main effect of Target Type [F (2, 118) = 425.00, $p < .001$, $\eta^2 = 0.88$], with accuracy for *Scrambled* Targets (M = 37.63%, SE = 1.72%) significantly lower than *Left* (M = 60.02%, SE = 1.67%) and *Right* Targets (M = 61.98%, SE = 1.38%). The interaction effect between Set Size and Target Type was also significant [F (6, 354) = 67.73, $p < .001$, $\eta^2 = 0.53$].

Simple main effects revealed significant differences across Set Size at each Target Type. For *Left* Targets [F (3, 531) = 3.58, $p = .014$, $\eta^2 = 0.02$], accuracy was significantly lower for Four Left Targets than Two Left Targets, but no other differences were found (One Left, M = 60.94%, SE = 2.06%; Two Left, M = 63.19%, SE = 1.49%; Three Left, M = 60.68%, SE = 1.14%; Four Left, M = 55.28%, SE = 2.00%). For *Right* Targets [F (3, 1.49) = 13.46, $p < .001$, $\eta^2 = 0.07$], accuracy was significantly higher for One Right Target compared to other Set Sizes. Accuracy was also lower for Two compared to Four Right Targets (One Right, M = 70.46%, SE = 1.61%; Two Right, M = 54.8%, SE = 1.44%; Three Right, M = 60.15%, SE = 1.13%; Four Right, M = 62.52%, SE = 1.35%). Accuracy for *Scrambled* Targets [F (3, 1.49) = 131.43, $p < .001$, $\eta^2 = 0.43$] decreased significantly as Set Size increased (One Scrambled, M = 66.03%, SE = 2.23%; Two Scrambled, M = 37.49%, SE = 1.67%; Three Scrambled, M = 27.59%, SE = 1.52%; Four Scrambled, M = 19.4%, SE = 1.46%).

Simple main effects also found significant differences between Target Types at each Set Size. At Set Size One [F (2, 472) = 13.05, $p < .001$, $\eta^2 = 0.05$], accuracy was highest for *Right*, then *Scrambled*, followed by *Left* Targets. At Set Size Two [F (2, 472) = 98.73, $p < .001$, $\eta^2 = 0.29$], accuracy was highest for *Left*, *Right*, and then *Scrambled* Targets. At Set Size Three [F (2, 472) = 206.57, $p < .001$, $\eta^2 = 0.47$], accuracy for *Left* and *Right* Targets was not significantly different and both conditions were higher in accuracy than *Scrambled* Targets. At Set Size Four [F (2, 472) = 306.51, $p < .001$, $\eta^2 = 0.56$], accuracy was highest for *Right*, followed by *Left*, and then Scrambled Targets.

A.3.5.2 Detection Reaction Time

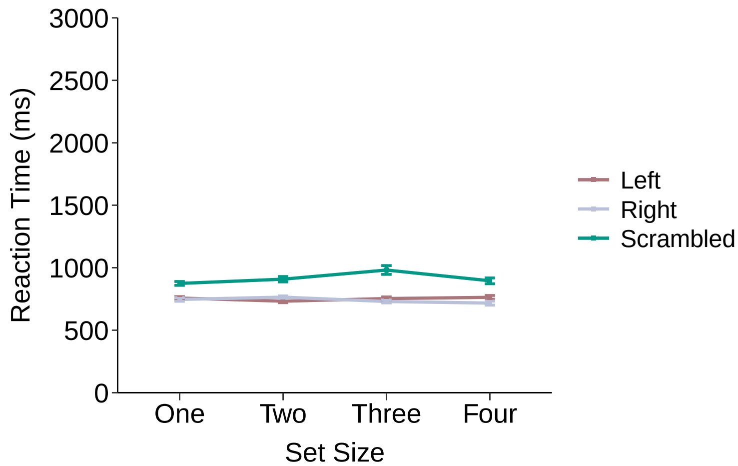

Figure A.16 Mean detection RT for Left, Right, and Scrambled conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.16 reports RT data for Experiment 12 for each Target Type at each Set Size. To investigate detection RT, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Left, Right, Scrambled*). The analysis revealed no significant main effect of Set Size [F (3, 177) = 1.55, p = .203, η^2 = 0.03], with similar RTs for *One* (M = 792 ms, SE = 14 ms), *Two* (M = 801 ms, SE = 15 ms), *Three* (M = 822 ms, SE = 20 ms), and *Four* Targets (M = 792 ms, SE = 18 ms). However, there was a significant main effect of Target Type [F (2, 118) = 81.87, p < .001, η^2 = 0.58], with similarly fast RTs for *Left* (M = 751 ms, SE = 13 ms) and *Right* Targets (M = 739 ms, SE = 13 ms) compared to *Scrambled* Targets (M = 915 ms, SE = 24 ms). A significant interaction effect between Set Size and Target Type was also found [F (6, 354) = 3.80, p = .001, η^2 = 0.06].

Simple main effects revealed no significant differences across Set Sizes for either *Left* [F (3, 531) = 0.56, p = .640, η^2 = 0.00] or *Right* conditions [F (3, 1.49) = 1.32, p = .268, η^2 = 0.01] (One Left, M = 757 ms, SE = 11 ms; Two Left, M = 732 ms, SE = 13 ms; Three Left, M = 753 ms, SE = 13 ms; Four Left, M = 763 ms, SE =

14 ms; One Right, M = 746 ms, SE = 14 ms; Two Right, M = 764 ms, SE = 10 ms; Three Right, M = 730 ms, SE = 12 ms; Four Right, M = 717 ms, SE = 16 ms). However a significant difference was found at the *Scrambled* condition [$F(3, 1.49) = 6.86$, $p < .001$, $\eta^2 = 0.04$] where RTs slowed from One, to Two, to Three Scrambled Targets, and then became slightly faster at Four Scrambled Targets (One Scrambled , M = 874 ms, SE = 15 ms; Two Scrambled , M = 908 ms, SE = 21 ms; Three Scrambled , M = 982 ms, SE = 35 ms; Four Scrambled , M = 895 ms, SE = 23 ms).

Simple main effects for Target Type at each Set Size revealed the same pattern. *Left* and *Right* Targets were not significantly different from each other but were significantly faster than *Scrambled* Targets [Set Size One, $F(2, 472) = 14.98$, $p < .001$, $\eta^2 = 0.06$; Set Size Two, $F(2, 472) = 26.14$, $p < .001$, $\eta^2 = 0.10$; Set Size Three, $F(2, 472) = 57.73$, $p < .001$, $\eta^2 = 0.20$; Set Size Four, $F(2, 472) = 25.48$, $p < .001$, $\eta^2 = 0.10$].

A.3.6 Experiment 13: Faces and Sideways in Real Scenes (250 ms)

A.3.6.1 Detection Accuracy

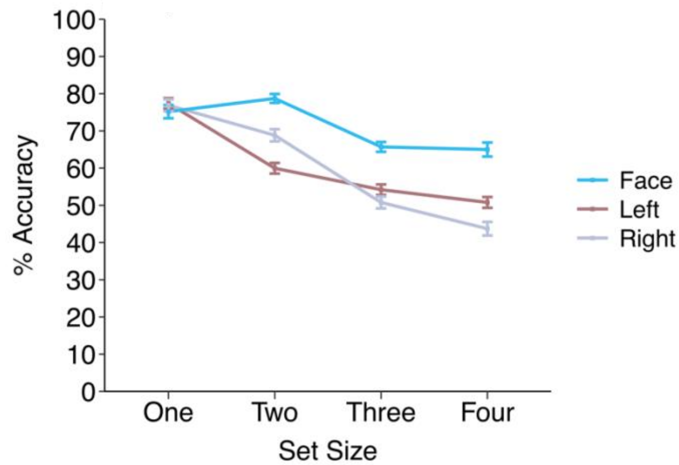

Figure A.17 Mean percentage detection accuracy for Left, Right, and Scrambled conditions at each Set Size.Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.17 reports accuracy data for Experiment 13 for each Target Type at each Set Size. To investigate detection accuracy, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright*, *Left, Right*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 66.07, $p < .001$, $\eta^2 = 0.53$], with the accuracy decreasing as Set Size increased (*One* , M = 76.48%, SE = 1.59%; *Two* , M = 69.16%, SE = 1.42%; *Three* , M = 56.89%, SE = 1.44%; *Four* , M = 53.16%, SE = 1.72%). There was also a significant main effect of Target Type [F (2, 118) = 84.11, $p < .001$, $\eta^2 = 0.59$], with accuracy for *Upright* Targets (M = 71.14%, SE = 1.54%) significantly higher than *Left* (M = 60.58%, SE = 1.42%) and *Right* Targets (M = 60.05%, SE = 1.67%). The interaction effect between Set Size and Target Type was also significant [F (6, 354) = 23.50, $p < .001$, $\eta^2 = 0.28$].

Simple main effects revealed significant differences across Set Size at each Target Type. For the *Upright* condition [F (3, 531) = 16.66, $p < .001$, $\eta^2 = 0.09$], accuracy was similarly high at One and Two Upright Targets compared to Three and Four Upright Targets which did not differ in accuracy (One Upright, M = 75.17%, SE = 1.77%; Two Upright, M = 78.72%, SE = 1.2%; Three Upright, M = 65.7%, SE = 1.32%; Four Upright, M = 64.99%, SE = 1.88%). For the *Left* condition [$F(3, 531) = 49.62$, $p < .001$, $\eta^2 = 0.22$], accuracy was significantly higher at One Left Target compared to the other condition, accuracy for Two Left Targets was also higher than Four Left Targets (One Left, M = 77.39%, SE = 1.40%; Two Left, M = 59.96%, SE = 1.43%; Three Left, M = 54.2%, SE = 1.41%; Four Left, M = 50.78%, SE = 1.45%). For the *Right* condition [$F(3, 531) = 84.26$, $p < .001$, $\eta^2 = 0.32$], accuracy decreased as Set Size increased (One Right, M = 76.88%, SE = 1.61%; Two Right, M = 68.82%, SE = 1.64%; Three Right, M = 50.77%, SE = 1.6%; Four Right, M = 43.72%, SE = 1.83%).

Simple main effects also revealed significant differences between Target Types at each Set Size, except Set Size One [$F(2, 472) = 0.82$, $p = .441$, $\eta^2 = 0.00$]. Across Set Size Two [$F(2, 472) = 53.39$, $p < .001$, $\eta^2 = 0.18$], accuracy was highest for *Upright*, followed by *Right*, and then *Left* Targets. At Set Size Three [$F(2, 472) = 37.11$, $p = .001$, $\eta 2 = 0.14$], accuracy for *Upright* was significantly higher than *Left* and *Right* which did not differ from each other. At Set Size Four [$F(2, 472) = 71.15$, $p < .001$, $\eta^2 = 0.23$], accuracy was highest for *Upright*, followed by *Left*, and then *Right* Targets.

A.3.6.2 Detection Reaction Time

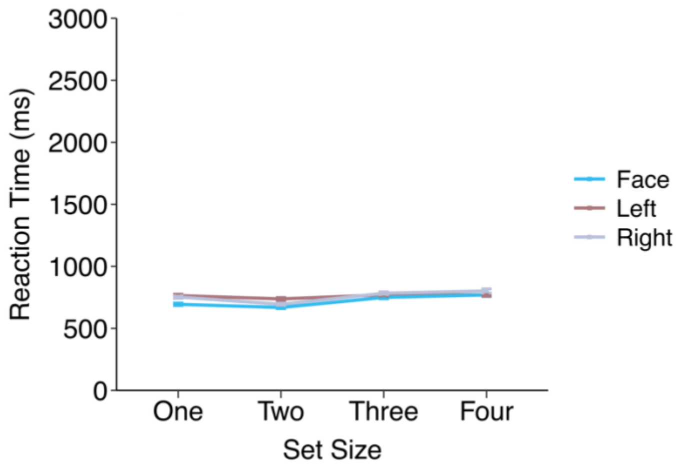

Figure A.18 Mean detection RT for Upright, Left, and Right conditions at each Set Size. Error bars show within-subjects standard error (Cousineau, 2005).

Figure A.18 reports RT data for Experiment 13 for each Target Type at each Set Size. To investigate detection RTs, the data were subjected to a two-way ANOVA with within-subjects factors of Set Size (*One, Two, Three, Four*) and Target Type (*Upright*, *Left, Right*). The analysis revealed a significant main effect of Set Size [F (3, 177) = 14.79, $p < .001$, $\eta^2 = 0.20$], with intermediate RT for *One* (M = 737, SE = 11), followed by faster RT at *Two* (M = 700, SE = 10), that increased at *Three* (M = 768, SE = 11), and *Four* Targets (M = 780, SE = 13). There was also a significant main effect of Target Type [F (2, 118) = 20.81, $p < .001$, $\eta^2 = 0.26$], with similarly fast RTs for *Upright* Targets (M = 721, SE = 11) compared to *Left* (M = 760, SE = 11) and *Right* Targets (M = 758, SE = 12). A significant interaction effect between Set Size and Target Type was also found [F (6, 354) = 4.09, $p = .001$, $\eta^2 = 0.06$].

Simple main effects revealed significant differences across Set Sizes for *Upright* [F (3, 531) = 14.34, $p < .001$, $\eta^2 = 0.07$] and *Right* conditions [F (3, 1.2) = 14.28, $p < .001$, $\eta^2 = 0.07$], but not the *Left* condition [F (3, 1.2) = 1.55, $p = .200$, $\eta^2 = 0.01$]. At the *Upright* condition, RT was similarly faster at One and Two Upright Targets compared to Three and Four Upright Targets which did not differ in from each other (One Upright, M = 694 ms, SE = 10 ms; Two Upright, M = 669 ms, SE =

9 ms; Three Upright, M = 750 ms, SE = 11 ms; Four Upright, M = 770 ms, SE = 13 ms). At the *Right* condition, RT was slowest at Four Right Targets, whilst RTs for One Right and Three Right Targets did not differ from each other (One Right, M = 753 ms, SE = 11 ms; Two Right, M = 694 ms, SE = 9 ms; Three Right, M = 783 ms, SE = 11 ms; Four Right, M = 802 ms, SE = 17 ms). RTs at the *Left* condition were similar (One Left, M = 764 ms, SE = 11 ms; Two Left, M = 737 ms, SE = 11 ms; Three Left, M = 770 ms, SE = 12 ms; Four Left, M = 769 ms, SE = 11 ms).

Simple main effects also compared between Target Types at each Set Size. Significant differences were found at Set Size One [F (2, 472) = 14, $p < .001$, η^2 = 0.06], where *Upright* Targets were detected the fastest. At Set Size Two [F (2, 472) = 11.96, p = .001, η2 = 0.05], RT for *Left* Targets was slowest compared to *Upright* and *Right* Targets. No significant differences were found at Set Size Three [F (2, 472) = 2.81, $p = .061$, η^2= 0.01]. A significant simple main effect was reported at Set Size Four [F (2, 472) = 3.40, $p = .034$, η^2= 0.01], but further Tukey's HSD tests revealed no differences.

References

Almazawi, A. (2013, May 5). *Waseem And Amjad Portraits*. Flickr.Com. https://www.flickr.com/photos/malim92/8539439500/

Alvarez, G. A., & Cavanagh, P. (2004). The Capacity of Visual Short-Term Memory Is Set Both by Visual Information Load and by Number of Objects. *Psychological Science*, *15*(2), 106–111. https://doi.org/10.1111/j.0963-7214.2004.01502006.x

Amso, D., Haas, S., & Markant, J. (2014). An Eye Tracking Investigation of Developmental Change in Bottom-up Attention Orienting to Faces in Cluttered Natural Scenes. *PLOS ONE*, *9*(1), e85701. https://doi.org/10.1371/JOURNAL.PONE.0085701

Anobile, G., Castaldi, E., Moscoso, P. A. M., Burr, D. C., & Arrighi, R. (2020). "Groupitizing": a strategy for numerosity estimation. *Scientific Reports*, *10*(1). https://doi.org/10.1038/S41598-020-68111-1

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Ashleigh, C. (2012, March 11). *Ngatho Mugo an AYAD talks with her colleagues at ICDDR,B who are researching maternal child health at the Centre for Child and Adolescent Health*. Flickr.Com. https://www.flickr.com/photos/106853342@N04/10710842256

Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201–221. https://doi.org/10.1016/S0010-0277(02)00178-6

Bindemann, M., & Burton, A. M. (2008). Attention to upside-down faces: An exception to the inversion effect. *Vision Research*, *48*(25), 2555–2561. https://doi.org/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. https://doi.org/10.1111/j.1551-6709.2009.01035.x

Bindemann, M., Burton, A. M., Hooge, I. T. C., Jenkins, R., & De Haan, E. H. F. (2005). Faces retain attention. In *Psychonomic Bulletin and Review* (Vol. 12, Issue 6, pp. 1048–1053). Psychonomic Society Inc. https://doi.org/10.3758/BF03206442

Bindemann, M., Jenkins, R., & Burton, A. M. (2007). A bottleneck in face identification: Repetition priming from flanker images. *Experimental Psychology*, *54*(3), 192–201. https://doi.org/10.1027/1618-3169.54.3.192

Bindemann, M., & Lewis, M. B. (2013a). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin and Review*, *20*(6), 1140–1145. https://doi.org/10.3758/s13423-013-0445-9

Bindemann, M., & Lewis, M. B. (2013b). Face detection differs from categorization: Evidence from visual search in natural scenes. *Psychonomic Bulletin and Review*, *20*(6), 1140–1145. https://doi.org/10.3758/S13423-013-0445-9/TABLES/2

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

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Brown, V., Huey, D., & Findlay, J. M. (1997). Face detection in peripheral vision: Do faces pop out? *Perception*, *26*(12), 1555–1570. https://doi.org/10.1068/P261555

Bruce, V., & Young, A. (2013). Face perception. *Face Perception*, 1–496. https://doi.org/10.4324/9780203721254/FACE-PERCEPTION-VICKI-BRUCE-ANDY-YOUNG

Bruckmaier, M., Tachtsidis, I., Phan, P., & Lavie, N. (2020). Attention and Capacity Limits in Perception: A Cellular Metabolism Account. *Journal of Neuroscience*, *40*(35), 6801–6811. https://doi.org/10.1523/JNEUROSCI.2368-19.2020

Bruyer, R., & Brysbaert, M. (2011). Combining speed and accuracy in cognitive psychology: Is the inverse efficiency score (IES) a better dependent variable than the mean reaction time (RT) and the percentage of errors (PE)? *Psychologica Belgica*, *51*(1), 5–13. https://doi.org/10.5334/pb-51-1-5

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

Calder, A. J. (2011). Does facial identity and facial expression recognition involve separate visual routes? *Oxford Handbook of Face Perception*. https://doi.org/10.1093/OXFORDHB/9780199559053.013.0022

Cavanagh, P., & Alvarez, G. A. (2005). Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, *9*(7), 349–354. https://doi.org/10.1016/J.TICS.2005.05.009

Chi, M. T. H., & Klahr, D. (1975). Span and rate of apprehension in children and adults. *Journal of Experimental Child Psychology*, *19*(3), 434–439. https://doi.org/10.1016/0022-0965(75)90072-7

Choo, H., & Franconeri, S. L. (2014). Enumeration of small collections violates Weber's law. *Psychonomic Bulletin and Review*, *21*(1), 93–99.

https://doi.org/10.3758/s13423-013-0474-4

Corkum, V., & Moore, C. (1998). The origins of joint visual attention in infants. *Developmental Psychology*, *34*(1), 28–38. https://doi.org/10.1037/0012-1649.34.1.28

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology*, *1*(1), 42–45. https://doi.org/10.20982/TQMP.01.1.P042

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114. https://doi.org/10.1017/S0140525X01003922

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision*, *10*(4), 1–17. https://doi.org/10.1167/10.4.16

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

Dakin, S. C., & Watt, R. J. (2009a). Biological "bar codes" in human faces. *Journal of Vision*, *9*(4), 2–2. https://doi.org/10.1167/9.4.2

Dakin, S. C., & Watt, R. J. (2009b). Biological "bar codes" in human faces. *Journal of Vision*, *9*(4), 2–2. https://doi.org/10.1167/9.4.2

De Marco, D., & Cutini, S. (2020). Introducing CUSTOM: A customized, ultraprecise, standardization-oriented, multipurpose algorithm for generating nonsymbolic number stimuli. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01332-z

Demeyere, N., Rotshtein, P., & Humphreys, G. W. (2012). The neuroanatomy of visual enumeration: Differentiating necessary neural correlates for subitizing versus counting in a neuropsychological voxel-based morphometry study. *Journal of Cognitive Neuroscience*, *24*(4), 948–964. https://doi.org/10.1162/jocn_a_00188

Devue, C., Belopolsky, A. V., & Theeuwes, J. (2012). Oculomotor Guidance and Capture by Irrelevant Faces. *PLOS ONE*, *7*(4), e34598. https://doi.org/10.1371/JOURNAL.PONE.0034598

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. https://doi.org/10.1016/J.JECP.2012.04.012

Doty, T. J., Japee, S., Ingvar, M., & Ungerleider, L. G. (2013). Fearful Face Detection Sensitivity in Healthy Adults Correlates with Anxiety-Related Traits. *Emotion (Washington, D.C.)*, *13*(2), 183. https://doi.org/10.1037/A0031373

Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology*, *6*(5), 178–190. https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8

Dunbar, R. I. M. (2012). The social brain meets neuroimaging. *Trends in Cognitive Sciences*, *16*(2), 101–102. https://doi.org/10.1016/j.tics.2011.11.013

Dunbar, R. I. M., & Spoors, M. (1995). Social networks, support cliques, and kinship. *Human Nature*, *6*(3), 273–290. https://doi.org/10.1007/BF02734142

Duncan, J., & Humphreys, G. W. (1989). Visual Search and Stimulus Similarity The Efficiency of Visual Selection. In *Psychological Review* (Vol. 96, Issue 3).

Eayrs, J., & Lavie, N. (2018). Establishing individual differences in perceptual capacity. *Journal of Experimental Psychology: Human Perception and*

Performance. https://doi.org/10.1037/xhp0000530

Eayrs, J. O., & Lavie, N. (2021a). Perceptual load and enumeration: Distractor interference depends on subitizing capacity. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(9), 1149. https://doi.org/10.1037/XHP0000823

Eayrs, J. O., & Lavie, N. (2021b). *Perceptual Load and Enumeration: Distractor Interference Depends on Subitizing Capacity*. Journal of Experimental Psychology: Human Perception and Performance. https://doi.org/10.1037/xhp0000823

Egeth, H., Kwak, H., Green, B., Mul-lin, P., Pringle, R., Sejnowski, T., & Yuhas, B. (1988). *Subitizing: Direct apprehension or serial processing?*

Farah, M. J., Tanaka, J. W., & Drain, H. M. (1995). What Causes the Face Inversion Effect? *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 628–634. https://doi.org/10.1037/0096-1523.21.3.628

Farah, M. J., Wilson, K. D., Tanaka, J. N., & Drain, M. (1998). What is “special” about face perception? In *Human Perception: Institutional Performance and Reform in Australia* (Vol. 105, Issue 3). https://doi.org/10.4324/9781351156288-26

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(14), 9602–9605. https://doi.org/10.1073/pnas.152159999

Farroni, T., Johnson, M. H., Menon, E., Zulian, L., Faraguna, D., & Csibra, G. (2005). Newborns’ preference for face-relevant stimuli: Effects of contrast polarity. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(47), 17245–17250. https://doi.org/10.1073/pnas.0502205102

Farroni, T., Menon, E., Rigato, S., & Johnson, M. H. (2007). The perception of facial expressions in newborns. *Https://Doi.Org/10.1080/17405620601046832*, *4*(1), 2–13. https://doi.org/10.1080/17405620601046832

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. In *Trends in Cognitive Sciences* (Vol. 8, Issue 7, pp. 307–314). Elsevier Current Trends. https://doi.org/10.1016/j.tics.2004.05.002

Frank, M. C., Vul, E., & Johnson, S. P. (2009). Development of infants' attention to faces during the first year. *Cognition*, *110*(2), 160. https://doi.org/10.1016/J.COGNITION.2008.11.010

Freiwald, W., Duchaine, B., & Yovel, G. (2016). FACE PROCESSING SYSTEMS: FROM NEURONS TO REAL WORLD SOCIAL PERCEPTION. *Annual Review of Neuroscience*, *39*, 325. https://doi.org/10.1146/ANNUREV-NEURO-070815-013934

Fysh, M. C. (2018). Individual differences in the detection, matching and memory of faces. *Cognitive Research: Principles and Implications*, *3*(1), 20. https://doi.org/10.1186/s41235-018-0111-x

Garner, K. G., Bowman, H., & Raymond, J. E. (2021). Incentive value and spatial certainty combine additively to determine visual priorities. *Attention, Perception, and Psychophysics*, *83*(1), 173–186. https://doi.org/10.3758/S13414-020-02124-W

Garrido, L., Duchaine, B., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology*, *2*(Pt 1), 119–140. https://doi.org/10.1348/174866407X246843

Gebuis, T., Cohen Kadosh, R., & Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: A critical review. *Acta Psychologica*, *171*, 17–35.

https://doi.org/10.1016/j.actpsy.2016.09.003

Gilmore, C., Attridge, N., & Inglis, M. (2011). Measuring the approximate number system. *Quarterly Journal of Experimental Psychology*, *64*(11), 2099–2109. https://doi.org/10.1080/17470218.2011.574710

Gilmore, C., Göbel, S. M., & Inglis, M. (2018a). An Introduction to Mathematical Cognition. In *An Introduction to Mathematical Cognition*. https://doi.org/10.4324/9781315684758

Gilmore, C., Göbel, S. M., & Inglis, M. (2018b). Nonsymbolic Number. In *An Introduction to Mathematical Cognition* (pp. 6–28). Routledge. https://doi.org/10.4324/9781315684758-2

Goffaux, V., & Dakin, S. C. (2010). Horizontal information drives the behavioral signatures of face processing. *Frontiers in Psychology*, *0*(SEP), 143. https://doi.org/10.3389/FPSYG.2010.00143/XML/NLM

Guillaume, C., Guillery-Girard, B., Chaby, L., Lebreton, K., Hugueville, L., Eustache, F., & Fiori, N. (2009). The time course of repetition effects for familiar faces and objects: An ERP study. *Brain Research*, *1248*, 149–161. https://doi.org/10.1016/J.BRAINRES.2008.10.069

Guillaume, M., Nys, J., Mussolin, C., & Content, A. (2013). Differences in the acuity of the Approximate Number System in adults: The effect of mathematical ability. *Acta Psychologica*, *144*(3), 506–512. https://doi.org/10.1016/j.actpsy.2013.09.001

Hjelmås, E., & Low, B. K. (2001). Face detection: A survey. *Computer Vision and Image Understanding*, *83*(3), 236–274. https://doi.org/10.1006/cviu.2001.0921

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

https://doi.org/10.1068/p3252

Honey, C., Kirchner, H., & VanRullen, R. (2008). Faces in the cloud: Fourier power spectrum biases ultrarapid face detection. *Journal of Vision*, *8*(12), 9–9. https://doi.org/10.1167/8.12.9

Infusionsoft. (2009, August 7). *A Sales Gang with Michael Gerber*. Flickr.Com. https://www.flickr.com/photos/infusionsoft/3799939948/in/photolist-6MMFL1

Inglis, M., & Gilmore, C. (2013). Sampling from the mental number line: How are approximate number system representations formed? *Cognition*, *129*(1), 63–69. https://doi.org/10.1016/j.cognition.2013.06.003

Izard, V., & Dehaene, S. (2008). Calibrating the mental number line. *Cognition*, *106*(3), 1221–1247. https://doi.org/10.1016/j.cognition.2007.06.004

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

Kamaludin, K. Z. (2011, February 1). *Family vacation: Shanghai*. Flickr.Com. https://www.flickr.com/photos/muslimpage/5531069770/

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *Journal of Neuroscience*, *17*(11), 4302–4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Karras et al., & Nvidia. (2018). *Generated Photos | Unique, worry-free model photos*. https://generated.photos/

Katzin, N., Cohen, Z. Z., & Henik, A. (2019). If it looks, sounds, or feels like subitizing, is it subitizing? A modulated definition of subitizing. *Psychonomic Bulletin and Review*, *26*(3), 790–797. https://doi.org/10.3758/s13423-018-1556-

0

Kauffmann, L., Khazaz, S., Peyrin, C., & Guyader, N. (2021). Isolated face features are sufficient to elicit ultra-rapid and involuntary orienting responses toward faces. *Journal of Vision*, *21*(2), 1–24. https://doi.org/10.1167/jov.21.2.4

Kaufman, E. L., & Lord, M. W. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498–525. https://doi.org/10.2307/1418556

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. https://doi.org/10.1111/desc.12829

Kelly, D. J., Quinn, P. C., Slater, A. M., Lee, K., Gibson, A., Smith, M., Ge, L., & Pascalis, O. (2005). Three-month-olds, but not newborns, prefer own-race faces. *Developmental Science*, *8*(6), F31. https://doi.org/10.1111/J.1467-7687.2005.0434A.X

Keys, R. T., Taubert, J., & Wardle, S. G. (2021). A visual search advantage for illusory faces in objects. *Attention, Perception, & Psychophysics 2021 83:5*, *83*(5), 1942–1953. https://doi.org/10.3758/S13414-021-02267-4

Kingstone, A., Tipper, C., Ristic, J., & Ngan, E. (2004). The eyes have it!: An fMRI investigation. *Brain and Cognition*, *55*(2), 269–271. https://doi.org/10.1016/J.BANDC.2004.02.037

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1–16. https://doi.org/10.2/JQUERY.MIN.JS

Kramer, R. S. S., Jenkins, R., & Burton, A. M. (2017). InterFace: A software package for face image warping, averaging, and principal components analysis. *Behavior*

Research Methods, *49*(6), 2002–2011. https://doi.org/10.3758/s13428-016-0837-7

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. https://doi.org/10.1068/p230095

Kumar, A., Kaur, A., & Kumar, M. (2018). Face detection techniques: a review. *Artificial Intelligence Review 2018 52:2*, *52*(2), 927–948. https://doi.org/10.1007/S10462-018-9650-2

Laidlaw, K. E. W., Badiudeen, T. A., Zhu, M. J. H., & Kingstone, A. (2015). A fresh look at saccadic trajectories and task irrelevant stimuli: Social relevance matters. *Vision Research*, *111*(Part A), 82–90. https://doi.org/10.1016/J.VISRES.2015.03.024

Langton, S. R. H., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition*, *107*(1), 330–342. https://doi.org/10.1016/J.COGNITION.2007.07.012

Lavie, N. (1995). Perceptual Load as a Necessary Condition for Selective Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(3), 451–468. https://doi.org/10.1037/0096-1523.21.3.451

Lavie, N., & De Fockert, J. W. (2003). Contrasting effects of sensory limits and capacity limits in visual selective attention. *Perception & Psychophysics 2003 65:2*, *65*(2), 202–212. https://doi.org/10.3758/BF03194795

Le Grand, R., Cooper, P. A., Mondloch, C. J., Lewis, T. L., Sagiv, N., de Gelder, B., & Maurer, D. (2006). What aspects of face processing are impaired in developmental prosopagnosia? *Brain and Cognition*, *61*(2), 139–158. https://doi.org/10.1016/j.bandc.2005.11.005

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

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

Lewis, M. B., & Ellis, H. D. (2003). How we detect a face: A survey of psychological evidence. *International Journal of Imaging Systems and Technology*, *13*(1), 3–7. https://doi.org/10.1002/ima.10040

Liu, J., Harris, A., & Kanwisher, N. (2002). Stages of processing in face perception: An MEG study. *Nature Neuroscience*, *5*(9), 910–916. https://doi.org/10.1038/nn909

Lyons, I. M., & Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition*, *121*(2), 256–261. https://doi.org/10.1016/j.cognition.2011.07.009

MacEntee, S. (2010, September 4). *Home Office*. Flickr.Com. https://www.flickr.com/photos/112494386@N08/12085578295

Mazza, V., & Caramazza, A. (2015). Multiple object individuation and subitizing in enumeration: A view from electrophysiology. *Frontiers in Human Neuroscience*, *9*(APRIL), 162. https://doi.org/10.3389/FNHUM.2015.00162/XML/NLM

McGrath, J. E., Arrow, H., & Berdahl, J. L. (2000). The study of groups: Past, present, and future. *Personality and Social Psychology Review*, *4*(1), 95–105. https://doi.org/10.1207/S15327957PSPR0401_8

Ming, M. (2014, January 12). *Migrant Worker*. Flickr.Com. https://www.flickr.com/photos/112494386@N08/12085578295

Murofushi, K. (1997). Numerical Matching Behavior by a Chimpanzee (*Pan troglodytes*): Subitizing and Analogue Magnitude Estimation. *Japanese Psychological Research*, *39*(3), 140–153. https://doi.org/10.1111/1468-5884.00050

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(42), 16598–16603. https://doi.org/10.1073/pnas.0703913104

Nojiri, K. (2011, December 1). *Wedding Photos*. Flickr.Com. https://www.flickr.com/photos/79992579@N07/8471363651

Norman, D. A., & Bobrow, D. G. (1975). On data-limited and resource-limited processes. *Cognitive Psychology*, *7*(1), 44–64. https://doi.org/10.1016/0010-0285(75)90004-3

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

O'Toole, A. J. (2005). Psychological and Neural Perspectives on Human Face Recognition. In *Handbook of Face Recognition*. https://doi.org/10.1007/0-387-27257-7_16

O'Toole, A. J., & Castillo, C. D. (2021). Face Recognition by Humans and Machines: Three Fundamental Advances from Deep Learning. *Https://Doi.Org/10.1146/Annurev-Vision-093019-111701*, *7*, 543–570. https://doi.org/10.1146/ANNUREV-VISION-093019-111701

Omer, Y., Sapir, R., Hatuka, Y., & Yovel, G. (2019). What Is a Face? Critical Features for Face Detection. *Perception*, *48*(5), 437–446. https://doi.org/10.1177/0301006619838734

Pascalis, O., De Haan, M., Nelson, C. A., & De Schonen, S. (1998). Long-term recognition memory for faces assessed by visual paired comparison in 3- and 6-month-old infants. *Journal of Experimental Psychology: Learning Memory and Cognition*, *24*(1), 249–260. https://doi.org/10.1037/0278-7393.24.1.249

Piazza, M., Fumarola, A., Chinello, A., & Melcher, D. (2011). Subitizing reflects visuo-spatial object individuation capacity. *Cognition*, *121*(1), 147–153. https://doi.org/10.1016/j.cognition.2011.05.007

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

Prunty, J. E., Jenkins, R., Qarooni, R., & Bindemann, M. (2022). Ingroup and outgroup differences in face detection. *British Journal of Psychology (London, England : 1953)*. https://doi.org/10.1111/BJOP.12588

Puce, A., McNeely, M. E., Berrebi, M. E., Thompson, J. C., Hardee, J., & Brefczynski-Lewis, J. (2013). Multiple faces elicit augmented neural activity. *Frontiers in Human Neuroscience*, *7*(MAY), 282. https://doi.org/10.3389/fnhum.2013.00282

Purcell, D. G., & Stewart, A. L. (1986). *The face-detection effect*.

Purcell, D. G., & Stewart, A. L. (1988). The face-detection effect: Configuration enhances detection. In *Perception & Psychophysics* (Vol. 43, Issue 4). https://doi.org/10.3758/BF03208806

Pylyshyn, Z. W. (2004). Some puzzling findings in multiple object tracking: I. Tracking without keeping track of object identities. *Visual Cognition*, *11*(7), 801–822. https://doi.org/10.1080/13506280344000518

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

Radford, C., McNutt, J. W., Rogers, T., Maslen, B., & Jordan, N. (2020). Artificial eyespots on cattle reduce predation by large carnivores. *Communications Biology*, *3*(1), 1–8. https://doi.org/10.1038/s42003-020-01156-0

Railo, H., Karhu, V. M., Mast, J., Pesonen, H., & Koivisto, M. (2016). Rapid and accurate processing of multiple objects in briefly presented scenes. *Journal of Vision*, *16*(3), 8–8. https://doi.org/10.1167/16.3.8

Railo, H., Koivisto, M., Revonsuo, A., & Hannula, M. M. (2008). The role of attention in subitizing. *Cognition*, *107*(1), 82–104. https://doi.org/10.1016/J.COGNITION.2007.08.004

Riggs, K. J., Ferrand, L., Lancelin, D., Fryziel, L., Dumur, G., & Simpson, A. (2006). Subitizing in tactile perception. *Psychological Science*, *17*(4), 271–272. https://doi.org/10.1111/j.1467-9280.2006.01696.x

Robertson, D. J., Jenkins, R., & Burton, A. M. (2017). Face detection dissociates from face identification. *Visual Cognition*, *25*(7–8), 740–748. https://doi.org/10.1080/13506285.2017.1327465

Roos, T. (1993). Voronoi diagrams over dynamic scenes. *Discrete Applied Mathematics*, *43*(3), 243–259. https://doi.org/10.1016/0166-218X(93)90115-5

Roper, Z. J. J., Cosman, J. D., & Vecera, S. P. (2013). Perceptual load corresponds with factors known to influence visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1340–1351. https://doi.org/10.1037/A0031616

Rosburg, T., Ludowig, E., Du¨mpelmann, M., Du¨mpelmann, D., Alba-Ferrara, L., Urbach, H., & Elger, C. E. (2009). The effect of face inversion on intracranial and scalp recordings of event-related potentials. *Wiley Online Library*, *47*(1), 147–157. https://doi.org/10.1111/j.1469-8986.2009.00881.x

Rossion, B., Joyce, C. A., Cottrell, G. W., & Tarr, M. J. (2003). Early lateralization and orientation tuning for face, word, and object processing in the visual cortex. *NeuroImage*, *20*(3), 1609–1624. https://doi.org/10.1016/J.NEUROIMAGE.2003.07.010

Scaife, M., & Bruner, J. S. (1975). The capacity for joint visual attention in the infant. *Nature*, *253*(5489), 265–266. https://doi.org/10.1038/253265A0

Schauder, K. B., Park, W. J., Tsank, Y., Eckstein, M. P., Tadin, D., & Bennetto, L. (2019). Initial eye gaze to faces and its functional consequence on face identification abilities in autism spectrum disorder. *Journal of Neurodevelopmental Disorders*, *11*(1). https://doi.org/10.1186/S11689-019-9303-Z

Simion, F., Turati, C., Valenza, E., & Leo, I. (2005). The emergence of cognitive specialization in infancy : The case of face preference. *Researchgate.Net*, 189–208. https://www.researchgate.net/profile/Francesca-Simion/publication/281827933_The_emergence_of_cognitive_specialization_in_infancy_The_case_of_face_preference/links/55f9d89108aec948c497e2f1/The-emergence-of-cognitive-specialization-in-infancy-The-case-of-fa

Simpson, E. A., Buchin, Z., Werner, K., Worrell, R., & Jakobsen, K. V. (2014). Finding faces among faces: human faces are located more quickly and accurately than other primate and mammal faces. *Attention, Perception, and Psychophysics*, *76*(8), 2175–2183. https://doi.org/10.3758/s13414-014-0744-x

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. https://doi.org/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. https://doi.org/10.1016/j.jecp.2019.05.001

Simpson, E. A., Maylott, S. E., Mitsven, S. G., Zeng, G., & Jakobsen, K. V. (2019). Face detection in 2- to 6-month-old infants is influenced by gaze direction and species. *Developmental Science*, *23*(2). https://doi.org/10.1111/desc.12902

Sinha, P. (2002). Qualitative Representations for Recognition. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *2525*, 249–262. https://doi.org/10.1007/3-540-36181-2_25/COVER

Sinha, P., Balas, B., Ostrovsky, Y., & Russell, R. (2006). Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, *94*(11), 1948–1961. https://doi.org/10.1109/JPROC.2006.884093

Skelhorn, J., Holmes, G. G., Hossie, T. J., & Sherratt, T. N. (2016). Eyespots. *Current Biology*, *26*(2), R52–R54. https://doi.org/10.1016/j.cub.2015.10.024

Starkey, G. S., & McCandliss, B. D. (2014). The emergence of “groupitizing” in children’s numerical cognition. *Journal of Experimental Child Psychology*, *126*, 120–137. https://doi.org/10.1016/J.JECP.2014.03.006

Stein, T., Sterzer, P., & Peelen, M. V. (2012). Privileged detection of conspecifics: Evidence from inversion effects during continuous flash suppression. *Cognition*, *125*(1), 64–79. https://doi.org/10.1016/j.cognition.2012.06.005

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B.

(2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior Research Methods*, *48*(3), 1197–1204. https://doi.org/10.3758/s13428-015-0641-9

Tanaka, J. W., & Farah, M. J. (2007). Parts and wholes in face recognition. *Http://Dx.Doi.Org/10.1080/14640749308401045*, *46*(2), 225–245. https://doi.org/10.1080/14640749308401045

Thoma, V., & Lavie, N. (2013). Perceptual load effects on processing distractor faces indicate face-specific capacity limits. *Visual Cognition*, *21*(8), 1053–1076. https://doi.org/10.1080/13506285.2013.853717

Tomasello, M. (2019). The Cultural Origins of Human Cognition. In *The Cultural Origins of Human Cognition*. https://doi.org/10.2307/j.ctvjsf4jc

Townsend, J. T., & Ashby, F. G. (1978). Methods of Modeling Capacity in Simple Processing Systems. In *Cognitive theory (Volume 3)* (pp. 199–239). Psychology Press. https://doi.org/10.4324/9781315802473-14

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*(1), 97–136. https://doi.org/10.1016/0010-0285(80)90005-5

Trick, L. M. (1992). A theory of enumeration that grows out of a general theory of vision: Subitizing, counting, and finsts. *Advances in Psychology*, *91*(C), 257–299. https://doi.org/10.1016/S0166-4115(08)60889-4

Trick, L. M., & Pylyshyn, Z. W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80–102. https://doi.org/10.1037/0033-295X.101.1.80

Tsao, D. Y., & Livingstone, M. S. (2008). Mechanisms of Face Perception. *Annual*

Review of Neuroscience, *31*(1), 411–437. https://doi.org/10.1146/annurev.neuro.30.051606.094238

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

Valentine, T. (1991). A Unified Account of the Effects of Distinctiveness, Inversion, and Race in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A*, *43*(2), 161–204. https://doi.org/10.1080/14640749108400966

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Vandierendonck, A. (2018). Further tests of the utility of integrated speed-accuracy measures in task switching. *Journal of Cognition*, *1*(1). https://doi.org/10.5334/joc.6

Vandierendonck, A. (2020). On the utility of integrated speed-accuracy measures when speed-accuracy trade-off is present. *Journal of Cognition*, *4*(1), 1–26. https://doi.org/10.5334/joc.154

Vestner, T., Tipper, S. P., Hartley, T., Over, H., & Rueschemeyer, S. A. (2019). Bound together: Social binding leads to faster processing, spatial distortion, and enhanced memory of interacting partners. *Journal of Experimental Psychology. General*, *148*(7), 1251–1268. https://doi.org/10.1037/XGE0000545

Vuilleumier, P. O., & Rafal, R. D. (2000). A systematic study of visual extinction Between-and within-field deficits of attention in hemispatial neglect. In *Brain* (Vol. 123).

Wang, R., & He, X. (2019). Face detection based on template matching and neural network. *Proceedings - 2019 International Conference on Communications, Information System, and Computer Engineering, CISCE 2019*, 547–550. https://doi.org/10.1109/CISCE.2019.00127

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), 1–14. https://doi.org/10.1038/s41467-020-18325-8

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238. https://doi.org/10.3758/BF03200774

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience 2004 5:6*, *5*(6), 495–501. https://doi.org/10.1038/nrn1411

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from backgrounds in visual search tasks. *Vision Research*, *42*(28), 2985–3004. https://doi.org/10.1016/S0042-6989(02)00388-7

Xu, X., & Biederman, I. (2014). Neural correlates of face detection. *Cerebral Cortex*, *24*(6), 1555–1564. https://doi.org/10.1093/cercor/bht005

Yin, R. K. (1969a). Looking at upide-down faces. *Journal of Experimental Psychology*, *81*(1), 141–145. https://doi.org/10.1037/H0027474

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

Yovel, G., & Kanwisher, N. (2005a). The neural basis of the behavioral face-inversion effect. *Current Biology*, *15*(24), 2256–2262. https://doi.org/10.1016/j.cub.2005.10.072

Yovel, G., & Kanwisher, N. (2005b). The Neural Basis of the Behavioral Face-Inversion Effect. *Current Biology*, *15*(24), 2256–2262. https://doi.org/10.1016/J.CUB.2005.10.072

Zhou, W.-X., Sornette, D., Hill, R. A., & Dunbar, R. I. M. (2005). Discrete hierarchical organization of social group sizes. *Proc. R. Soc. B*, *272*, 439–444. https://doi.org/10.1098/rspb.2004.2970