The Visual Processing of Human Faces and Bodies as Visual Stimuli in Natural Scenes.

By

Victoria Ruth Kroll

A thesis submitted in partial fulfilment of the requirements of Nottingham Trent University for the degree of Doctor of Philosophy.

April 2016

Abstract

How faces are recognized and detected has been the focus of an extensive corpus of research. As such, it is now well established that human faces can be detected rapidly in a visual scene and that they automatically capture a viewer's attention over other objects. However, under natural viewing conditions the human face is attached to a substantial cue, the human body. The evidence to-date of a similar attentional processing advantage for human bodies is less clear. This is remarkable given the social significance and evidence of neural specificity for these stimuli. Additionally, most previous investigations of preferential attention towards faces and bodies have presented these stimuli in simple displays, namely uniform colour backgrounds (Bindemann, Scheepers, Ferguson & Burton, 2010). Therefore, this thesis aimed to address the relationship between attention and face and body processing in natural scenes directly by assessing the consequences of numerous experimental manipulations in both a visual search paradigm and additional singleton paradigm. The first line of enquiry examined participants' ability to detect face and body stimuli in comparison to other objects in natural scenes. Subsequent experiments examined whether faces and bodies captured attention when they were task-irrelevant. In line with previous research, the main findings indicate that human faces do have attentional advantages and capture attention in both natural and grey scenes. They also indicate that human bodies (without the head) do not have detection advantages over other objects, nor do they capture attention in a bottom-up manner. Any biases or detection advantages observed for body targets are because they larger in size than other objects or because they are odd stimuli in that scene. Human full-body targets (including the face), which are perceived on a day-to-day basis, capture attention partly because they include face and partly because they are large objects in the scene. These findings modify claims of person perception suggesting that the detection of a full-body in natural scenes is facilitated by attention capture by faces, any advantages from bodies are the result of attention capture by their large size, rather than some attentional advantage. Future investigating into face and body processing should use natural backgrounds to gain a more realistic insight in to face and body processing in the real world.

Acknowledgements

First and foremost I would like to thank Dr Andrew Dunn for his support and excellent supervision throughout this project. He has always been there with a shoulder to cry on and to help me through the good and bad times, of which there has been plenty during the past 5 years. Whether it has been some thesis related stress/disaster or personal crisis he has always made time to lend his ear and provide the occasional cocktail. I owe him a massive thank you for getting me to the stage I am at today and pushing me through times when I did not believe that I could do it. Thanks are also due to Christina Howard and Thom Baguley for their guidance and input.

I would also like to take this opportunity to thank my parents, Margaret and Geoffrey Kroll. They have supported me all the way through my university years, listening to me moan/cry/scream at all stages from my degree to thesis, always believing in me. Their unwavering support and love has got me through this project, without them I have no idea where I would be today.

I would also like to thank my best friend Karima Susi. We started this journey together on day one of Masters, where we both meekly told one another we were on the 1+3 bursary and our friendship has grown ever since. Her intelligence, dedication and motivation have pushed me to work harder than I ever thought I could. Without her loyalty, support and many a crazy night out to forget/drown our sorrows; I could not have done this.

I would also like to give a special thank you to two of the most important people in my life, Ruth Kroll and Margaret Nycz. Both had an enormous impact on my life especially when I was growing up and I genuinely would not be the person I am today if it were not for both of them.

There are too many other people that I need to thank for their support throughout the creation of this thesis, but I cannot fit them all in here. So a massive thank you goes out to everybody else who has helped me during the course of the thesis.

Contents

1 Table of Contents

Table of Figures

Chapter 1

General Introduction

1 Literature Review

Selective attention is the ability to select and attend visually to certain aspects of the environment. This is often done in preference over other aspects. Much is known about selective visual attention (e.g., Pashler, 1998; Styles, 1997) and it is now generally accepted that certain stimuli can automatically capture a viewer's attention (e.g., Abrams & Christ, 2006; Gellantly & Cole, 2000; Remington, Johnson & Yantis, 1992). One such stimulus is the human face. A large amount of behavioural evidence suggests that human faces can be detected rapidly in a visual scene and automatically capture a viewer's attention (Bindemann & Burton, 2009; Bindemann, Burton, Langton, Schweinberger, & Doherty, 2007; Langton, Law, Burton, & Schweinberger, 2008; Ro, Friggel, & Lavie, 2007; Ro, Russell, & Lavie, 2001; Theeuwes & Van der Stigchel, 2006). However, whilst the evidence for attention capture by faces is extensive, there is comparatively less behavioural evidence in favour of a similar attentional advantage for human bodies. This is remarkable as faces and bodies are rarely perceived without the other and that there are arguably no other two stimuli that are so equally matched in terms of their social and biological significance. However, most previous investigations of preferential attention towards faces and bodies have presented these stimuli in simple displays, namely uniform colour backgrounds (Bindemann et al., 2010). As a result, any findings of advantages for faces and bodies observed in these simple backgrounds are confined to these simple displays and are therefore limited in what they can reveal about attention capture by faces and bodies in the real world (Bindemann et al., 2010).

The present literature review aims to discuss what is currently known about the visual processing of human faces and bodies as visual stimuli and establish the research question for the present thesis. It begins with an introduction to the relative principles of attention. The current behavioural literature on faces and bodies is then discussed to establish what is understood about the attentional processing of these stimuli. Related to this, relevant neuroscientific evidence will be discussed in order to provide a comprehensive review of the processing of face and body stimuli. The aim of the review is to highlight that like faces, human bodies may also exhibit similar attentional processing advantages.

1.1 Principles of Attention

The visual system is presented with a complex information rich complex environment. Indeed, it is estimated that the amount of information that is transferred from the retina to the brain is in the range of 10^8–10^9 bits per second (Deco, Pollatos, & Zihl, 2002). The limited high-level representation capacity processing of the human brain along with a combination of other factors, e.g., small fovea; frontal eye positions and small visual fields, restricts the number of objects and locations that can be fixated at any one point (Chun & Marois, 2002). In essence, the human brain simply does not have enough capacity to process all of this information at the same time and assimilate this into consciousness (e.g. Tsotsos, 1990).

To avoid this problem, the visual system has a number of ways of coping with the large input of information. First, the visual system discards input, in that rather than high-level processing occurring at all locations, the best resolution is confined to the fovea and therefore only information here is processed in full (Wolfe, 1994; Wolfe & Horowitz, 2004). In the periphery, information is much more crudely processed with large losses in acuity occurring only a few degrees into it (Wolfe & Horowitz, 2004). However, reducing the input of information on its own is not sufficient to solve this problem.

Second, the visual system also copes by a number of attentional mechanisms that select of a small set of stimuli to process, whilst subjecting the rest to minimal processing. For example, in a visual field it is possible to read any text, however, it is not possible to read two texts at two different locations in a visual field at the same time (Wolfe, 1994). It is however, widely agreed that visual processes are themselves are not limited to one part of the visual field at any one time (Wolfe, 1994). Rather, analysis of the world takes place in two consecutive stages: a pre-attentive stage; i.e. before selection occurs, this operates in parallel over large portions of the visual field (e.g. Neisser, 1967), and a second stage of capacity limited processes that produces a more detailed analysis of the visual scene and serial in its deployment from one location to another. This process of shifting the focus of attention is known as focal or selective attention, for the purpose of this thesis it will be referred to as selective attention (Broadbent, 1958; Kahneman, 1973; Neisser, 1967). The concept of this dichotomy of visual processing has been incorporated into many models of attention, which are discussed in depth later in this chapter. However, the locus of selection, whether attention limits occur before the stimulus is identified (early selection theories) or after (late selection theories), is a long-standing issue.

Early selection theories as proposed originally by Broadbent (1958) and further developed by Treisman (1969; Treisman & Geffen, 1967) and others (e.g., Kahneman, 1973; Neisser, 1967) posit that selection is necessary for full perceptual processing. Importantly, it is argued that perception is a limited capacity and that selection occurs at an early, pre-categorical level based on the physical attributes of the stimuli, such as colour, shape, or spatial location. In this view, a preliminary analysis of the physical attributes operates in parallel over the visual field. A selective filter set for these physical attributes whereby selective attention only allows information that fits with these attributes to pass through the filter to the limited capacity channel to be subsequently identified. As a result, unattended stimuli are not processed beyond their initial physical attributes (Broadbent, 1958). Information then enters the senses where it is then held in a preattentive store. Subsequent research rendered early selection theory questionable. For example, in one well-known experiment, Moray (1959) found that during a shadowing task, where participants repeated words back that were presented in an attended ear and ignored words in an unattended ear, participants we able to detect their own name in the unattended ear. Therefore, unattended information had not been fully excluded from further analysis, as would be predicted by early selection theory. This led to an opposing late selection theory. Originally developed by Deutsch and Deutsch (1963) and further by others (e.g., Norman, 1968; Duncan, 1980; Tipper, 1985), the late selection theory argues that before attentional selection can occur, identification of all input occurs in parallel across the visual field. In this view, identity of both attended and unattended stimuli are computed in a capacity-unlimited manner without the need for selection. Selection accordingly begins after the entire visual field has been analysed, with attentional limitations arising from the selection of which of the identified stimuli will gain access to cognitive representations, memory and responses (e.g. Deutsch & Deutsch, 1963, Duncan, 1980, Norman, 1968). The debate between late and early theories of attention is therefore an issue of the level to which unattended stimuli are processed preattentively (Lamy & Tsal, 2001). Attempts to resolve this debate, however, showed just as much evidence for late selection (e.g. Yantis & Johnson, 1990), as for early selection (e.g. Pashler, 1984). Lavie (1995) offered perceptual load theory as a compromise for this debate. Central to this theory is that early or late selection is conditional on the perceptual load of the relevant information within the visual field. The main benefit of this theory is that it combines assumptions from both early and late selection theories. Specifically, it argues that attentional processing is limited capacity (early selection theory) and that processing is automatic in the way that it cannot be voluntarily withheld (late selection theory; Driver, 2001). Within this perceptual processing is assumed to occur automatically until the available capacity is exhausted. Load theory therefore predicts that under conditions of high perceptual load, when many relevant stimuli are present, attentional resources will be exhausted and early selection occurs. However, under low perceptual load if there is any spare capacity after the relevant stimuli, excess attentional resources will unintentionally capture spare capacity. In support of her theory, Lavie (1995) examined the effect of task-irrelevant distractor letters, under several conditions of varying perceptual load. In line with the theory, distractor inference depended on the task load, with distractor inference effects only being present in low-load conditions.

1.1.1 Visual Search: Feature Integration Theory

Although Lavie's perceptual load theory argues in favour of a compromise between early and late selection, it is an inflexible account of perception, namely, distractors are either processed or they are not. In contrast, the beginning point for most contemporary theories of selective attention is Feature Integration Theory (FIT), which argues that all stimuli, task relevant and irrelevant, will receive processing, even if just at the simple feature level and not identified for further selection (e.g., Treisman & Gelade, 1980). Unlike the notion of a selective filter, FIT mostly focuses on the role of attention in selecting and binding information. One source of support for FIT comes from visual search tasks. Indeed, the visual search paradigm has been one of the key paradigms over the last three decades in attention research examining theories of selective attention (Müller & Krummenacher, 2006). The following section will describe the visual search paradigm and in turn discuss FIT in more detail. In the typical visual search paradigm, participants are instructed to search for a specified target in a display, with a variable number of distractors. The number of items in the display is referred to as the set-size. The target may or not may be displayed in the visual display; however, it is the participant's task to make the target-present or target-absent decision as fast and accurately as possible (Treisman & Gelade, 1980). Typically, the average reaction time (RT) is plotted as a function of the total number of distractors present within the scene. The slope of the function gives an approximation of the search time cost for each additional distractor item in the display, measured in terms of time per display item. Early investigations of RT experiments showed an apparent dichotomy in the proficiency with which attentional processes can detect a target (Chan & Hayward, 2009; Treisman & Gelade, 1980; Treisman, Sykes & Gelade, 1977). Based on this, two modes of visual search were proposed: parallel and serial search (Treisman & Gelade, 1980). These corresponded to Neisser's (1967) description of preattentive and attentive processes. Parallel search is defined by flat search slopes, this essentially means that if the time taken to determine that a target is present is short and does not increase substantially with the increasing number of distractor items in the display (< 10ms/item), the target is detected in parallel with the visual system examining all items at the same time, without capacity limitations. This type of search is also considered to be automatic or preattentive, with the number of distractor items within a display having a relatively small effect on search time (Neisser, 1967; Wolfe, Cave & Franzel, 1989). This is also referred to as "pop out" and indicates that the target is summoning attention automatically. Conversely, if the time taken to detect a target increases linearly with number of distractors in an array (> 10ms/item), visual search is said to be serial, requiring sequential focused attention. Specifically, attention is directed to each item in the display one by one, allowing each item to be categorized as a target or a distractor sequentially (Sternberg, 1966 as cited in Wolfe, 2007). The increase in reaction times and ratio of the slopes (target present/target absent) of less than 2.0 indicates the presence of a serial, self-terminating search (Wolfe, 1998). Conversely, a ratio of more than 2.0 indicates parallel search.

Throughout the visual search research it is often noted that some search tasks are easy and efficient, in parallel, whereas others are inefficient, serial. For example, consider

Figure 1., the task of locating the blue target, the big target, and the leaning target can be done with relative ease and it is obvious that can be achieved simply regardless of the distractor set size. The defining attributes of these targets; i.e. colour, size and orientation guide attention efficiently to the target (Wolfe & Horowitz, 2004). Conversely, finding the 5 amongst the 2 is more difficult because spatial positioning information that makes these letters different (e.g. a 2 is a backwards 5) does not guide attention (Wolfe & Horowitz, 2004). The more distractor 2s there is in the scene, the harder the task will be. To explain this inconsistency, several models have been proposed. These models seek to explain how visual attention is deployed in order to detect a target (Cooper, Law & Langton, 2013), and which properties and attributes can guide attention and what cannot.

Figure 1. Examples of easy and difficult visual searches. It is easy to location the big ‘5’, the leaning ‘5’ and the red ‘5’. It is more difficult to located the ‘2’ within the display. From: Wolfe and Horowitz (2004) What attributes guide the deployment of visual attention and how do they do it?

As mentioned previously, the starting point, and probably most influential model of selective attention and visual search is Feature Integration Theory (FIT; Treisman & Gelade, 1980). FIT assumes that a visual scene is initially coded by separate dimensions such as colour, orientation, spatial frequency, brightness, and direction of movement. The term *dimension* is used to refer to the complete range of variation; for example, colour or orientation, and the term *feature* to refer to a particular value on a dimension; for example, yellow or horizontal, respectively. FIT argues that visual search is a two-stage model: the preattentive parallel stage and serial search.

In the parallel stage of search, only simple features can be detected with no attentional limits. This results in flat or shallow slopes, as number of distractors in the display has a minimal effect on reaction time to target detection. For instance, feature search occurs when the target differs from the distractors by the same elemental feature, such as colour: a pink target is found quickly if all distracters are green.

According to FIT, feature search is achievable by perceptual dimensions being represented in ranges of feature maps, with each feature represented by a separate map (e.g. red or green in the continuum of colour maps). Treisman (1982) argued that the separate feature maps record the presence of stimulus features independently, and that activity in a feature map indicates the presence of a feature (Quinlan, 2003). Therefore, if the target feature can be identified within a single dimension, it will be the only active feature map, facilitating pop out or attention capture by that feature (Chan & Hayward, 2009). However, not all targets are made up of one feature. Rather, many targets are a conjunction of many basic features and are not identifiable without taking into account multiple feature maps. The search for a conjunction target therefore requires the joining of two or more feature maps. A second, more comprehensive, cross-dimensional, self-terminating serial search is therefore assumed to occur.

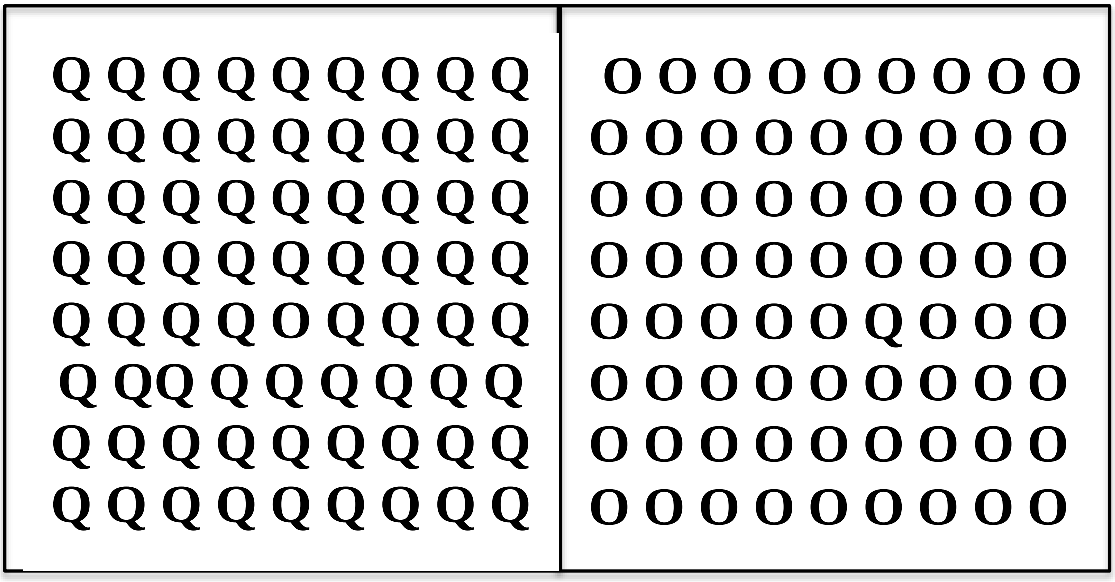
Central to the FIT is the notion that selective attention recombines or integrates these features in to a coherent object image from separate feature maps to guarantee the correct combination or conjunction of features for a target. Features that occur at the same location are combined to make a single object. According to early versions of FIT (Treisman 1985, 1986), this serial stage of the integration of features takes places with regards to a master map of locations. Separate feature maps are connected to the master map and represent particular points of salient activations in particular locations on the master map. Attention selects one location at a time on the master map and in doing this retrieves the features that are linked to the corresponding locations in a number of separable feature maps (Treisman, 1985: see Figure 2, Panel a).

Figure 2. Panel A illustrates serial attention locating a location on the master map, in doing this it retrieves all the feature information that is linked to that location. Panel B demonstrates Treisman and Sato (1990) revision showing the inhibition of active non-target locations on the master map. Panel C demonstrates guided search, showing how feature maps activate on the saliency map. From Dent et al., 2012, Parallel Distractor Rejection as a Binding Mechanism in Search.

Selective attention operates to integrate feature information conveyed by the separate feature maps. The serial nature of focused attention is assumed to be the spotlight of attention and is the glue forming the conjunctions from the available features. Selective attention is therefore the integration or feature-binding mechanism in this theory (Quinlan, 2003). The search of the master map will therefore take longer with an increase in number of items in an array. This is because there will be more salient activations on the master map, explaining why the time to detect feature combinations increases the more items/distractors there are in an array.

FIT has received heavy criticism when several studies demonstrated flat slopes for conjunction searches with two or more conjunctions (Nakayama & Silverman, 1986; Steinman, 1987; McLeod, Driver, & Crisp, 1988). As a result, FIT was revised to address this issue (Treisman & Sato, 1990). In the revision, top-down processes inhibit items with non-target features in parallel. Hence, when searching for red target amongst red and green targets, activation from the green targets will be inhibited. The master map of locations receives inhibition from the feature maps that contain the task-irrelevant features (Quinlan, 2003). This inhibition of active locations on the master map leads to a reduction in inference from distractors and a more efficient search for the target (See Figure 2 Panel b). As with the original FIT theory, the serial scan integrates relevant feature maps, however, conjunction search is made more efficient by the inhibition of non-target features (Quinlan, 2003).

Notably, one important factor of FIT is the notion of search asymmetries. This is the finding that search for the presence of an added feature that defines a target, is easier than the search for the absence of the feature (Treisman & Souther, 1985). For instance, Treisman and Souther (1985) showed that a Q pops out of an array of O’s (See Figure 3 Panel B), whereas an O in an array of Q’s does not (See Figure 3 Panel A). The critical feature here is the presence of the intersecting line on the O shape, making it a Q.



Panel A. Panel B.

Figure 3. Visual representation of a search asymmetry. As shown in Panel A it is harder to search for a O amongst Q’s than it is to search in Panel B for a Q among O’s.

Treisman and Souther (1985) argued that search for the target with the critical feature was in parallel, whereas the absence of the feature required a serial scan of all the items in the array. Importantly, FIT posits that search asymmetries can be accounted for by the existence of a feature map for coding the presence of a feature. The existence of such a search asymmetry is taken to indicate that that particular feature has a feature map specific for its automatic processing. Therefore, a search asymmetry in visual search can be used as a diagnostic for determining which features are detected pre-attentively in the visual system and automatically summon attention (Treisman & Gormican, 1988). This notion has been applied to other stimuli (e.g. faces, as discussed later; Hershler & Hochstein, 2005).

1.1.2 Guided Search

Despite many notable amendments to FIT, such as the inhibition of items with non-target features (Treisman & Sato, 1990), it remains a dichotomous approach, in which parallel processes do not inform following serial processes (Chan & Hayward, 2009). Specifically, FIT assumes that if a parallel search fails to locate a target, then there is no continuation of this information for serial search. Wolfe and colleagues (Wolfe, 1994; Wolfe et al., 1989) proposed Guided Search (GS) as an *alternative* to the feature integration theory. Wolfe et al. (1989) argued that the dichotomy of FIT made little sense. For instance, if a target was a blue Z amongst green Z's and blue O's, the target is a conjunction and so parallel processes cannot locate it. Parallel processes can however distinguish between the green and blue items in the display. Clearly, in this example no green item can be a blue Z. Based on examples like this, Wolfe et al. argued that it seems logical that parallel processes should be able to inform serial processes. Therefore in the examples like the one above, attention would not waste time examining the location of the green items as potential targets (Wolfe et al., 1989). In light of this, amending the FIT, GS theory of attention argues that parallel processes guide the spotlight of attention (selective attention) towards likely targets, hence the notion of guided search. Specifically, initial processing of a scene is assumed to occur in parallel throughout the visual field. Parallel representations are generated for a restricted amount of basic visual features (similar to FIT model of visual attention; Treisman, 1985), these are known as feature maps. For FIT, the parallel processing of these maps was to identify locations that were worthy of further attention. Within GS, this is modelled as a difference in activation of locations in each feature map. GS comprises that all items in a scene will be assigned an activation value, the greater the activation, the more likely it is that attention will be directed towards this item (see Figure 2 Panel C). There are two mechanisms that facilitate this activation: bottom-up activation and top-down activation. Bottom-up activation is determined solely by local visual properties of the scenes and is computed as the difference between and item and its neighbouring loci. For instance, a single red flower within a green field may draw attention more than the surrounding objects irrespective of the search target (Hwang, Higgins & Pomplun, 2009). Therefore bottom-up activation guides attention toward distinctive items in the field. Whereas, top-down user-driven activation results from features that describe a predefined target. The output of these two processes produces features maps with activation from both top-down and bottom-up activation. Then, the weighted sum of these activations forms an activation map, with the signal from each item indicating the likelihood of the target at that location. This activation does not reveal any information about its location, e.g. high activation from an orientation map is the same as high activation from the colour map. Attention is therefore guided by this activation map and directed toward the location of the highest activation first. If this location is not the target then attention will be redeployed to the areas of decreasing activation. To use the previous example, if the target is a blue Z, all the blue locations as well as the Z locations will be most excited on the activation map. Hence the blue Z location will be the most salient locations on the activation map and attention will be directed to this area. Thus, pop out, in the traditional sense, occurs when the activation value of a target is particularly high in relation to the items surrounding it in a scene.

As Wolfe et al. (1994; 1989) point out, if the passage of the information from parallel to serial processing for conjunctions were flawless, attention would be guided perfectly to the target every time, independent of distractor set size. However, serial search still occurs in visual search. According to GS theory this is due to noise signal being transferred from parallel to serial search. If this noise is large, the search will be serial. Conversely, if this noise is minimal, the target will be detected faster as incorrect targets will not be examined. The noise can be modelled as difference between participants and the use of different laboratories for testing (Wolfe et al., 1989). GS theory proposes that the information from parallel to serial is continuously being updated. Noise from the transmission from parallel to serial is at random, however, the signal is not. Therefore, at first, the incorrect target might be identified, however, over time the correct signal should emerge and the correct target identified. GS theory also provides a valid account of the observed flat slopes for targets defined by more than two conjunctions. For instance, the search for a conjunctively defined horizontal blue item benefits from concurrent top-down activation from the orientation map and the colour map. Therefore, there will be more activation at the target’s location, explaining efficient conjunction search performance. In sum, despite amendments to theories of visual search, GS remains a two-stage processing model, but assumes that the preattentive parallel stage of visual processing can guide serial search. However, importantly, taken together both FIT and GS theories of visual search provide a useful method for examining the processes that are involved in face detection.

1.1.3 Alternative Models of Visual Search

The classical theories of selective attention discussed thus far in this thesis suggest that there are two stages of visual processing involved in visual search: 1) a preattentive, parallel stage of processing where a limited set of features are detected in parallel (capacity unlimited) across the visual field and processed to extract a master or saliency map and 2) a serial (capacity limited) stage of processing within which selective attention is serially located to locations active or salient on the master or saliency map to identify items (Treisman & Gelade, 1980; Treisman, 1988; Wolfe, 1994, 2007). Consequently, both the classical FIT and GS theories of visual search are commonly referred to as two-stage serial models (Moran et al., 2015).

In contrast, alternative models of visual search have avoided the notion of the serial selection of items in a two-stage model by postulating a single parallel processing system (Muller & Krummenacher, 2006). In these models, primarily signal-detection-based, attention is distributed across the visual field and all items are identified simultaneously, with stimulus properties determining how fast an item can be identified (e.g., Eriksen & Spencer, 1969). However, highly contended is whether processing within parallel models is capacity limited (e.g. Ward & Mclellend, 1989; Snodgrass & Townsend, 1980), or unlimited (Palmer & Mclean, 1995).

Moreover, as noted in the literature, parallel models must be able to account for the linear increase in RTs with the increasing set-size, without resorting to explaining set-size increases in terms of a serial processing stage similar to that of FIT and GS (Muller & Krummenacher, 2006). Indeed, this pattern of data has been taken to be critical in distinguishing between serial or parallel models of visual search (Moran et al., 2015). In accordance with FIT, flat RT slopes have been interpreted as indicators of parallel search or pop out; conversely, steep slopes have been interpreted as an indicator of serial search processes (Treisman & Gelade, 1980). In line with GS, target salience determines whether it will be identified. With high target salience the target is identified fast, thus accounting for flat slopes in visual search. On the other hand, when target salience is low, the target is no more salient than distractors and item-selection is random and takes longer to find the target, accounting for steep slopes in visual search (Wolfe, 1994, 1998, 2007).

However, classical accounts of visual search have been challenged by supporters of parallel models, who suggest that parallel models can also account for increases in RTs as a function of set-size. As such, it has been argued that shallow slopes and steep slopes can be generated by both serial and parallel (capacity limited and unlimited) mechanisms, as explored in the following section (Thornton & Gilden, 2007; Townsend, 1972, 1976, 1990; Palmer & McLean, 1995; Palmer et al., 2000; Verghese, 2001; Ward & McClelland, 1989).

Both FIT and GS assume a high threshold whereby a distractor cannot be mistaken for a target. For instance, in FIT errors are generally ignored. An alternative method is to assume that distractors can be mistaken for targets. As such, many researchers have used Signal Detection Theory (SDT) to account for visual search. As such, in order to explore parallel models, it is first necessary to briefly explore SDT. In this framework, in visual search every item has an estimate of signal strength. An observer will make their target present/absent response, based on which has the large signal strength. If the maximum value is greater than some criteria the observer will give a target present response, else a target absent response is given (Palmer et al., 2011). Each added input (i.e., item) increases the probability of classifying a non-target as a target (i.e., false alarm). Errors can also occur when a target produces a signal that is misclassified as a distractor (i.e., a miss). As such, several parallel models of visual search have been based upon SDT.

Specifically, based on Ratcliff's (1978) *diffusion model*, derived from signal detection theory, Ward and McClelland (1989) suggest a capacity-limited parallel model that can account for steep slopes in visual search. In this model they argue that each item in a display has a detector that accumulates evidence for or against the hypothesis that it is the target of the search - so that search occurs in parallel over the visual field. Each detector has a positive, negative and resting threshold. Search terminates with a target present response, when a detector has crossed its positive threshold. Similarly, search terminates with a target absent response when all detectors have crossed their negative threshold (Ward & McClelland, 1989). The rate of accumulation is dependent on the amount of parallel resource available. Given a limited amount of processing resource an increase in the number of items in the visual scene will decrease the amount of resource per item. In turn, this will slow items journey to threshold and will produce steep slows with increasing set-size, similar to those in the serial stage of FIT and GS. As such, it is very difficult to distinguish between models containing a serial stage of processing and parallel models of visual search.

In addition, even unlimited-capacity parallel models of can account for set-size effects in terms of SDT. In line with SDT, Palmer and Mclean (1995) argue that if RTs are to remain the same with an increase in set-size, then the number of errors will increase. As such, the addition of items increases the noise and as a consequence observers are more likely to make a false alarm. Palmer and Mclean argue that in the event of a set-size increase, to maintain error rates at a reasonable and constant level, observers increase their decision criteria and slow down for larger set-sizes. In other words, in order to avoid false alarms, they raise their criteria for making a target present response. In turn, this slows down RT for making a correct target present response (i.e., hit). Within the literature a debate about capacity limits has existed for decades (Palmer et al., 2011). This is primarily because evidence for or against each model has been equal.

In sum, it appears that both serial and parallel models of visual search can account for set-size increases equivalently. In the literature more attention has been given to serial than to parallel models of visual search for various kinds of tasks. This is primarily because the equations for parallel models are difficult (Snodgrass & Townsend, 1980). Interestingly, a recent study compared a serial and parallel models' ability on three classic visual search tasks from a rich data set (Moran et al., 2015). They concluded that their results favour a two-stage serial model over a parallel model. However, the aim of this thesis is not to distinguish which type of model is the best fit for visual search, as such the parallel-serial debate shall not be discussed in any further detail.

1.1.4 Attentional Control: The Deployment of Attention

Given the enormous amount of information in visual scenes and the limited resources for visual processing, it is obvious the deployment of attention is not random (Wolfe, 1994). Rather, the deployment of attention, what is attended to and what is not under attentional control. It is now well established that two attentional mechanisms influence the control of visual attention within a visual scene: top-down or bottom-up attentional control (see, e.g., Eriksen & Hoffman, 1972; Posner, 1980; Theeuwes, 1991; Yantis & Jonides, 1990). Top-down or goal-directed control of attention is when an observer's intentions and goals influence voluntary or endogenous shifts of attention to stimuli that are selected for further visual analysis (Franconeri & Simons, 2003). For example, several studies have shown that prior knowledge/experience of a target's location leads to the target being responded to more efficiently and a reduction in the extent to which a distractor can capture attention (e.g. Posner, 1980). In contrast, bottom-up or stimulus-driven control of attention is when an object or event receives priority processing (e.g. Yantis, 1996). This process is driven exogenously, by an external stimulus, which automatically draws attention independent of a person's explicit goals and intentions (see section 1.2 for discussion on automaticity). This is often referred to attention capture and as a result, a viewer's attention may be reflexively drawn to a stimulus, even when it is contrary to the observer's wishes (French, Mareschal, Mermillod & Quinn, 2004; Franconeri & Simons, 2003; Bindemann et al., 2007). For example, the activation of the brake lights of the car in front of you in a traffic queue will draw attention automatically. There are in fact a large range of bottom-up visual attributes that can lead to attention capture (Chun & Wolfe, 2001). Indeed, bottom-up control of attention is often understood in terms of stimulus salience: the contrast between an object in a visual scene and its surroundings with regards to its low-level factors, such as orientation, texture and colour (Hershler & Hochstein, 2009). As such, bottom-up capture is also assumed to operate at an early visual stage, most probably at the level of feature search in FIT (Styles, 1997). In line with this, abrupt onset stimuli (and offset) have been shown to capture attention in an involuntary bottom-up or automatic manner (e.g., Theeuwes, 1991; Abrams & Christ, 2006). Evidence for this comes from demonstrations that attention is drawn to targets that were abruptly onset, even though responses to such targets were counterproductive to the task demands. For example, when a brief illumination (used as an abrupt visual onset) occurred at one location followed by a target at a different location, target detection was slow compared to a non-cued/valid cued location (e.g., Posner & Cohen, 1984). Conversely, responses were faster to valid pre-cued locations. Important here is that only on a small number of trials were the pre-cued locations predictive of the target location; it therefore would have been beneficial for participants' to ignore the cues as they were more often misleading as to the targets actual location. This suggests that participants' attention had been captured in an involuntary bottom-up manner by the cues, independently of their intentions.

A number of studies demonstrate similar attentional effects with more meaningful stimuli than basic colours, shapes, and onset/offset of stimuli. For instance, research has shown faster detection times for fear relevant stimuli (e.g., spiders and snakes) than fear-irrelevant stimuli (e.g., mushrooms and flowers), suggesting that some significant stimuli may have priority in the competition of selective attention (e.g., Ohman, Flykt & Esteves, 2001; Ohman & Mineka, 2001). However, few stimuli are as socially and biologically significant in the environment than the human face and body. This makes them prime candidates for automatic and preattentive processing (Palermo & Rhodes, 2007). While both face and body stimuli are visually dissimilar, they have many features in common. Both convey important information about a person, such as a person's identity, age, gender, intent and emotion (Bindemann & Burton, 2009). The failure to notice the presence of either a face or body in an environment would therefore lead to a loss of this information. As such, given that the ability to detect a face or body is a prerequisite to further processing, such as face and person recognition, the ability to rapidly detect a face or body not only represents one of the most important building blocks of human social cognition, but also one the most important aspects of person perception (Bindemann & Burton, 2009).

1.1.5 Human Faces and Bodies are Special

The specialness of the human face and body as visual stimuli is acknowledged by converging lines of both behavioural and neuroscientific evidence suggesting qualitatively different processes for face and body detection, as compared to object detection (Kanwisher, McDermott & Chun, 1997). The following sections will examine these lines of evidence in order to provide an account of the current literature of both stimuli and their preferential status within the visual system. It is important to note is that the processing of the human face and body is a multistage process. This ranges along a continuum from the ability to detect a face or body, the categorization, the identification and recognition of the presented face or body as being that of a specific individual (Hershler, Golan, Bentin & Hochstein, 2010). At this stage, it is useful to distinguish between face or body detection and other face or body processing stages. Face or body detection is an early stage of processing and involves the initial detection of the presence of a face or body in a scene. Whereas, face and body identification and recognition are later stages within visual processing, involving making distinctions between individuals (Slaughter, Stone & Reed, 2004). Fast detection of a face or body in a scene is a diagnostic for specialized attentional processing by these stimuli. The focus here will be on the detection or attention capture stage of processing, not other face or body processing stages, such as identification and recognition. With this in mind, the following review will examine whether faces and bodies receive enhanced, preattentive processing as described above. I shall begin by considering if faces are processed automatically/preattentively and subsequently whether they engage selective attention. The terms automatic and preattentive processing can be used interchangeably, however the remainder of the thesis will use the term automatic processing.

1.2 Are Faces Processed Automatically?

Posner and Snyder (1975a, 1975b) proposed three main characteristics that should be satisfied for a process to be classed as automatic. These are: it must not interfere with other on-going mental activities (i.e. it is capacity free); it must be without intention, such that it's processing is involuntary; and it must be without conscious awareness. In order to examine whether faces are automatically processed I shall discuss each of these criteria separately.

1.2.1 Capacity Free: Searching for Faces

The task of looking for a face in a scene is a visual search task. As a result, many researchers have applied a visual search paradigm to examine the ability of the human visual system to detect faces (Cooper et al., 2013). Important to note here is that both the FIT and GS interpretations of visual search provide a useful method for examining the processes that are involved in face detection. These accounts predict that if faces are processed automatically and capacity free, then they will be detected in parallel by feature search and pop out of a display, regardless of the number of distractors.

Early visual studies using simplistic faces reported that there was no pop out for faces. Nothdurft (1993) employed a visual search paradigm and ran a series of experiments (which Nothdurft referred to as series, rather than experiments) where the target could either be a human face, expression, or a control that contained the visual properties of a face, but did not produce the percept of a face e.g., the outline of a face with scrambled features. Targets were presented in arrays of up to 48 control items and the time taken to detect the target was measured. Search slopes were found to be considerably larger than 10ms per item in all series, except for series one, strongly suggesting that the search for face targets was serial. The exception to this was series 1 where shallow search (slopes <10ms) indicated parallel search. In this series, the face target had hair and was presented in an array of identical but inverted (upside down) distractors. To examine this further, they presented the same face targets but with the internal facial features removed. Again they observed shallow slopes, indicating that it was the visual cue of target hair that facilitated face pop out. Nothdurft subsequently concluded that faces (and expressions) are not detected in parallel and do not pop out of scenes.

In a related study, Kuehn and Jolicoeur (1994) examined the influence of the quality of the stimuli, the image orientation, and similarity of target stimuli on the visual search for faces. They again used schematic faces and found no advantage for upright face targets among inverted faces (slopes 50ms per item). The same result was also replicated when they used digitalized images of faces, which represented a better quality image of a face. Interestingly, although they concluded that faces do not pop out of an array of distractors, they did find that when the distractors were less face-like and did not contain any facial features, the slopes became shallower. This finding indicated that target detection may not have been as influenced by a set-size increase in non-face distractors than inverted face distractors, consistent with the notion of parallel search. However, a criticism of Northdurft's (1993) and Kuehn and Jolicoeur's (1994) research was that they only used schematic or digitalised representations of faces, which as Lewis and Ellis (2003) point out, is not a realistic representation of a face and cannot be generalised to face detection.

In another study, Brown, Huey and Findlay (1997) used black and white photographic face stimuli to examine whether faces pop out in a visual search task. Rather than measuring manual reaction time data, participants were instructed to move their eyes to target faces amongst inverted and jumbled face distractors. The time taken for participants to move their eyes to a target face was measured. Although no pop out for faces was observed for naïve participants, search performance did improve with participants who had had practice. The conclusion drawn was that faces do not pop out, further suggesting that faces are detected with serial and not parallel search (Treisman & Gelade, 1980).

Other research has focused on face valence and identity. Hansen and Hansen (1988) observed that the time taken to detect a happy face was longer in an array of angry faces than an angry face in an array of happy faces, suggesting that suggesting that the angry faces pop out of the array. However, Purcell, Stewart and Skov (1996) claimed that the pop out for these angry faces was due to a low-level confound on these stimuli; i.e. a dark spot on the chin of the angry faces. They replicated the study, but improved the stimuli (by removing the dark spot) and found that the angry faces no longer pop out of the happy face arrays.

1.2.1.1 Pop out for faces

All of the aforementioned studies either used abstract schematic images of faces, which may not be representative of a real face, or they used distractors that were either faces or objects that contained face-like properties (Lewis & Ellis, 2003). Importantly, under real world circumstances face detection is unlikely to occur in a context where a face is immediately surrounded by lots of other inverted faces (Lewis & Edmonds, 2005). Rather, faces are more likely to be surrounded by natural or domestic items. Given these reasons, it is an oversight to generalize their results to face detection on a whole.

In light of the findings reviewed above, Lewis and Edmonds (2005) used a visual search paradigm to conduct a series of experiments examining whether faces pop out of natural scenes and which properties of the face potentially facilitate face-pop out. The first experiment directly assessed whether faces pop out of natural scenes. They took sixteen still pictures of different characters from a popular television program and split each image up into nine rectangle cells, one of which contained a face, the remaining were background non-face information. They then constructed composites of the stills, resulting in square arrays with a varying number of rectangle cells (4, 9 and 16). On face-present trials, a face would be presented separately as one of these rectangle cells surrounded by background (non-face) cells in a square arrangement (array). Conversely, on face-absent trials none of the rectangle cells contained a face. They predicted that if faces pop out then an increase in the number of rectangle cells should not affect the time taken to detect a target face. Their results showed that the detection of a face was in fact slowed by an increase in the total number of rectangle cells in the display, however, this rate was lower than 10ms per item and slope for target absent trials was 15ms per item, giving a slope ratio of 5.0. Using visual search concepts, Lewis and Edmonds suggested that this demonstrates that, at least in scrambled scenes, faces are detected in parallel.

Given that this finding of pop out for faces was in direct contrast with previous research findings (Purcell et al., 1996; Nothduft, 1993; Kuehn & Jolicoeur, 1994; Hansen & Hansen, 1988), Lewis and Edmonds suggested that this discrepancy might be the result of the paradigm that they used. Traditional visual search tasks use targets surrounded by an array of distractors, however, they presented their stimuli in separate rectangle cells, which made up the arrays. They argued that this potentially led to the pop out observed for faces in their experiments. To explore this possibility they ensured that the paradigm they used could also produce serial search. For this purpose, they used the same paradigm, but they now presented inverted faces in the non-target cells in square arrays of 2, 9, and 16 cells. Using inverted faces replicated the type of distractors used in previous research. Similar to the previous literature, when inverted face distractors surrounded an upright target face, search slopes of 20ms per items for target present trials were observed. This yielded a slope ratio of less than 2.0, indicative of serial search.

Importantly, what this study demonstrates is that when a face is presented in face-like distractors, it does not pop out, however, surrounded by non-face scenic stimuli pop out for faces is observed. Lewis and Edmonds (2005) argued that using inverted faces as distractors essentially makes the paradigm an orientation-detection task, rather than a detection task and might explain the lack of pop out for faces in previous research (e.g., Nothdurft, 1993; Kuehn & Jolicoeur, 1994; Brown et al., 1999). Following their finding of pop out for faces, they then attempted to determine which features of the face facilitate pop out by looking at different transformations of their stimulus faces. Specifically, they examined face inversion, facial colour and Gaussian blur.

It is well established that inverting a face leads to a drop in accuracy and reaction times relative to upright faces (Yin, 1969; Haxby et al., 1999). This inversion effect has been attributed to the disruption of holistic processes that are used for processing upright faces (Farah, Tanaka & Drain, 1995; Tanaka & Farah, 1993). Specifically, faces are processed as a whole and not as separate parts (e.g., Farah, Wilson, Drain & Tanaka, 1998; McKone, Kanwisher & Duchaine, 2007). However, when a face is inverted the first order configuration of the face, i.e., the arrangement of features that define a face (e.g., eyes above the mouth), that makes holistic processing possible is disrupted (this is discussed in more detail in section 1.5). Importantly, first order configuration of the face is used to distinguish faces from other objects (face detection; Taubert, Apthorp, Aagtert-Murphy & Alais, 2010). Lewis and Edmond (2005) therefore argued that if holistic processing mediates pop out, then inverting the target faces will lead to steep search slopes. They also examined whether colour could facilitate pop out for faces, as faces may contain colours that make them more salient than their scenic distractors. To do this they compared face detection between greyscale and colour arrays. A Gaussian blur was also used to transform the face images. This transformation removes the high frequency details of the image giving the effect of it being far away. In turn, this removes the fine details of the image, which could potentially facilitate pop-up of faces. In addition they also they also applied luminance and hue reversal transformations to the faces. Previous research had demonstrated that luminance reversal impairs face processing (Kemp, Pike, White & Musselman, 1996). As Lewis and Edmonds (2005) note, reasoning for this may be because luminance reversal interferes with shading patterns that allow us to recognize a face shape. However, these findings were demonstrated with face recognition and not for face detection.

All slope gradients for each transformation were less than 10ms per item, indicative of pop out, except for the luminance-reversed faces. This allowed them to reject the possibility that colour, holistic processing, hue and the fine detail of the faces are pre-attentive guides for attention, responsible for the previously observed pop out. Luminance-reversed faces produced slope gradients that were consistent with serial search. They argued, in line with previous research, that this transformation might interfere with the clues that shading patterns provide to the shape of a face (Kemp et al., 1996). In sum, by transforming the images, Lewis and Edmonds (2005) were able to examine aspects of the face that are important for automatic processing. Their findings were inconsistent with previous claims that faces do not pop out of scenes, however they showed this might have been the result of unrealistic schematic face-stimuli and/or the use of distractors that were face-like. Accordingly, they provided some of the most robust evidence in favour of the automatic processing of faces, in ecologically valid contexts.

In a similar study, Hershler and Hochstein (2005) conducted a series of experiments, using the visual search paradigm to examine whether faces pop out. To test this assumption, they presented a variety of face and non-face (e.g., cars, houses) targets amongst a wide range of distractors from different object categories. Targets were presented on backgrounds of distractors that were homogenous (similar) in category.

Importantly, they demonstrated that faces were detected in parallel, whereas houses and cars were not. They also used the FIT concept of search asymmetry (Treisman & Souther, 1985: Refer to section 1.1.1), the easy detection of a target but not of its absence, to examine whether faces and objects pop out. Hershler and Hochstein observed a clear search asymmetry for faces: a target face was detected much easier on a background of distractor cars or houses, than the opposite search for a target car or house on a background of distractor faces. They also examined whether the features of the human face facilitate pop out for faces. To test this they scrambled the target faces, which despite disrupting the overall appearance of the face maintains the individual features of the face (i.e. eyes, nose, mouth). They argued if pop out for faces occurs because of a salient feature of the face when the faces were scrambled, pop out would still be observed. Scrambling the faces also provides a control for low-level image-based properties, such as edge density, colour, local contrast, which potentially make a face salient in a display and facilitate pop out (Langton et al., 2008). However, pop out was not observed for the scrambled faces, suggesting that pop out for faces is dependent on holistic information, rather than some individual feature of the face or some low-level image-based property.

Importantly, together the evidence from Hershler and Hochstein (2005) and from Lewis and Edmonds (2005) strongly indicate that so long as the target faces are sufficiently different from their distractors, faces will pop out of a scene. However, this conclusion has been controversial. Specifically, VanRullen (2006), after first demonstrating a pop out for faces, claimed that this effect can be attributed to low-level information, namely the Fourier amplitude spectrum. In order to examine this he used the inverse Fourier analysis method, in which the phase information of one image is kept the same, but the amplitude information is replaced by the amplitude information of another item. VanRullen (2006) made the assumption that the phase information of an image relates to the high-level processes, while the amplitude spectrum information relates to low-level processes. He found that when the amplitude spectrum was removed from the face images, pop out was no longer observed, leading to the conclusion that low-level amplitude factors facilitates pop out for faces.

In reply to VanRullen (2006), Hochstein and Hershlar (2006) argued that if his prediction is correct, then when the amplitude information from a face image is integrated with the phase information of another image and presented in an array of items that do not contain face amplitude information, the face should pop out again. However, this did not appear to be the case. This led to the conclusion that pop out for faces does not occur when the face target differs from the distractor only in its amplitude spectrum.

In summary, although much debate has surrounded whether faces pop out of displays, taken together the results from the reviewed visual search tasks appear to demonstrate that human faces do pop out of a visual display, regardless of how many other items are in the display. This is however contingent on faces being presented surrounded by distractors that are different to them. In turn, in line with Posner and Snyder (1975a, 1975b) criteria this suggests that faces are processed automatically, in the sense that they are processed capacity free.

1.2.2 Is Face Processing Mandatory?

Posner and Snyder (1975a, 1975b) criteria posit that for processing to be mandatory, it should happen outside of the viewer's intentions. Behavioural methods to examine whether face processing is mandatory include face-name interference tasks and inattention paradigms.

1.2.2.1 Face-Name Interference Tasks:

One of the biggest issues with the study of selective attention concerns the extent to which attention can prevent the perception of task irrelevant distractors. As previously discussed, load theory (Lavie, 1995; 2000) suggest that under conditions of low perceptual load (e.g. when just one stimulus is presented), distractors will be processed because any spare processing capacity will spill over to them and facilitate their perception. Conversely, under conditions of high perceptual load (e.g. when the task involves many relevant stimuli), all available capacity is exhausted and distractors will not be processed to perception (Lavie, 1995, 2000). As a result, experimental tasks that test load theory have been applied to the question of whether the perception of faces depends on the level of load in a relevant task (Lavie, Ro & Russell, 2003). If faces are processed regardless of the level of load then it would suggest that their processing is mandatory and unavoidable, and therefore automatic.

One of the most established behavioural techniques for measuring automatic processing is the target-distractor interference task (Bindemann, Burton & Jenkins, 2005a). In a typical target-distractor interference task, participants will be asked to make a response to a target, which could be presented with either a congruent distractor (the same as the target) or incongruent distractor (different from the target). For example, a participant may be asked to determine whether a face is male or female. A congruent distractor would be a female face, whereas an incongruent distractor would be a male face. Therefore, it can be inferred that the distractor was processed, if it can be shown that the reaction times to the target are slowed by an incongruent relative to congruent distractor.

A number of studies have demonstrated that distractor faces are processed when presented simultaneously with a non-face target (Jenkins, Burton, & Ellis, 2002; Lavie et al., 2003; Young, Ellis, Flude, McWeeney, & Hay, 1986). Young et al (1986) examined interference of concurrently presented photographs of faces and printed names of famous people in both a naming task and categorization task. During the naming task participants either stated the name of the face presented (either pop star or politician) or read aloud the printed name, in the presence or absence of a distractor. The distractor itself could either be the same occupation (congruent) or a different occupation (incongruent). Here they observed that naming faces in the congruent condition was faster than the naming of faces in the incongruent condition. More importantly, in a categorization task, where participants categorized whether the face or name belonged to the category of politicians or pop stars, they demonstrated that the presence of a face affected name categorization more than the presence of names affected the face categorization, strongly suggesting that the faces could not be ignored.

Extending this finding, Lavie et al. (2003) examined whether the level of attentional load affected the perception of distractor faces. In a similar experimental design, participants search for a name (politician or pop star) amongst a varying set size of letter strings (one, two, four, or six) in the presence of a distractor, which could be either congruent (face of person named) or incongruent (face of person from opposite occupation category). Like previous experiments that increase set size, the increase of set size had an effect on attentional demands. However, incongruent faces interfered with the categorization of the name at all set size levels, suggesting that participants were unable to ignore the distractor faces even under conditions of high perceptual load. They repeated this with familiar non-face distractors and found that incongruent non-face objects interfered with the task when the set size was four items. However, this effect was eliminated when the set size was increased to more than six items. In turn, Lavie et al. (2003) concluded that the processing of faces might be automatic and independent of attentional capacity.

1.2.2.2 Inattention Paradigms:

Information from unattended faces also seems to be processed. Evidence for this comes from a group of paradigms that elicit inattention phenomena. In such paradigms the failure to perceive a supra-threshold stimulus, a stimulus that is displayed just long enough to be perceived consciously, is said to be the consequence of attention being directed elsewhere or engaged by another task. One such example of this phenomenon is inattentional blindness (IB; Mack & Rock 1998). In a typical IB task, participants are engaged in a difficult task, such as judging the length of two lines on a target cross. After a number of trials, a task irrelevant stimulus (known as the critical stimulus) is also briefly displayed without the participants' prior knowledge. Participants are then asked if they noticed anything odd during the experiment. Remarkably, the standard finding is that a highly significant proportion of the participants report not seeing anything odd; i.e. they miss the critical stimulus, even though they are looking directly at it (e.g. Neisser, 1979). However, Mack & Rock (1998) found that when the critical stimulus was a schematic cartoon face, 85% of viewers detected a happy face, compared to 63% for a scrambled face, 40% for a frowning face and 15% for a simple geometric circle. This demonstrates that faces are more resistant to IB than other kinds of critical stimuli, suggesting that faces are appear to demand attention, even when not the target of attention.

Another example of automatic attention capture by faces can be seen in meta-contrasting masking (MM). MM tasks also indicate that the processing of faces is unavoidable. MM is the failure to notice a briefly flashed supra-threshold target that is followed (backward masking), or preceded (forward masking) by a masking stimulus.

Shelley-Tremblay and Mack (1999) used the notion of MM and presented schematic faces as both masks and targets. They observed that when the target was a face it was more likely to be noticed when it was masked compared to other stimuli. Importantly, when participants were not required to search for a target, they were asked post hoc whether they had seen anything other than the mask, they reported seeing a face more than other objects. The importance of this finding is that it demonstrates that even when faces are not explicitly searched for and are presented under conditions where their detection has been made very difficult, faces are detected and processed more than other objects.

Similar results have been found using the attentional blink (AB; Shapiro, 1994). The AB paradigm is the rapid serial visual presentation of a stream of stimuli, in which the participants' task is to locate two targets, generally referred to as the probe and target. Typically, the second target (the probe) is not noticed if it is presented between 150ms and 500ms after the first target (Raymond, Caldwell & Shapiro, 1992; Shapiro, 1994). This effect is said to be the result of attentional resources being occupied by the probe. Mack, Pappas, Silverman and Gay (2002) demonstrated that when a schematic smiley face was the target, between 92.9% and 87.3% of the time they were detected, however, tree shapes and inverted faces were repeatedly missed. Providing evidence that faces can be perceived under conditions in which other stimuli are not.

Taken together, the evidence from face-name interference and inattention paradigms are best understood in terms of mandatory processing of faces, whether they are the target of search or a distractor (Jenkins et al., 2002; Lavie, et al., 2003; Young, et al., 1986) and when their perception is made challenging (Raymond et al., 1992; Shapiro, 1994; Shelley-Tremblay & Mack, 1999; Mack & Rock, 1998; Mack et al., 2002).

1.2.3 Are Faces Processed Without Conscious Perception?

Not all stimuli that the brain processes are consciously perceived (Palermo & Rhodes, 2007; Rees, Kreiman, & Koch, 2002), as such, this type of processing is used to examine automatic processing (Morris, Pelphrey and McCarthy, 2007). Therefore, brain activity for stimuli that are not consciously perceived can be taken as an indicator that these stimuli are automatically processed without conscious awareness (Morris et al., 2007). In line with this, neuroimaging studies have shown that regions of the brain that are associated with face processing are activated by the presentation of faces that are outside of awareness. As will be discussed in more detail further in this introduction, the most robust findings for a face-selective module in the human brain are found on the lateral side of the mid-fusiform gyrus, known as the fusiform face area (FFA; e.g., Kanwisher et al., 1997; McCarthy et al., 1997; Puce, Allison, Gore, & McCarthy, 1995).

In a functional magnetic resonance (fMRI) experiment, Jiang and He (2006) presented participants with either neutral or fearful faces that were either visible or rendered invisible as a result of interocular suppression. They found that when the face stimuli were visible the FFA responded strongly. Interestingly, when the faces were rendered invisible activation in the FFA was also measured, though this activation was greatly reduced.

In a similar study, in order to examine whether faces presented outside of a participant's awareness can activate the FFA, Morris et al. (2007) measured participants' brain activation to masked, unmasked faces and non-face objects. To establish that their masking procedures were effective at preventing faces and non-face objects from entering conscious perception, participants first performed a behavioural task outside of the fMRI scanner. For this they were presented with masked faces and objects and were to give one response when the believed that they saw a face and another response when they saw an object. Target detection occurred on 8.2% of trials. For these detections, 52.7% (a value that did not differ from chance) of the time participants were able to accurately determine whether what they saw was a face or object. Morris et al. therefore concluded that even if participants were able to detect the target it was rare and when they did they were not able to determine what it was that they saw. In the scanner, participants performed a colour discrimination task, where they gave a response once they detected a coloured mosaic amongst a stream of greyscale mosaics. Unbeknown to the participants masked faces and objects were also shown every 12 – 18 seconds. Participants also viewed an unmasked version of the experiment to localize and identify activations in the FFA that could be compared to activations evoked by masked stimuli. They observed that unmasked faces engaged the FFA. However, importantly, they observed that masked faces engaged the FFA significantly more compared to the masked objects. These results therefore suggest that the processing of faces in the FFA is automatic.

Research using neglect and extinction patients, who are cortically blind because of damage to the right inferior parietal lobe, has also provided an understanding of the processing of stimuli that are not consciously perceived. Neglect is characterised by patients seemingly ignoring people, object and events presented in the visual field contralesional (predominately left) to the damaged hemisphere (predominately right;

Heilman, Watson & Valenstein, 1993). Whereas patients with extinction may lack awareness of a stimulus presented on the contralesional side of the visual field to the damaged hemisphere when a stimulus is presented simultaneously in the ipsilesional visual field of the damaged hemisphere. Interestingly, in these patients, the visual pathways to the occipital and temporal cortex are often intact (Vuilleumier et al., 2002). Both behavioural and neuroimaging data indicate that despite the patients unawareness of stimuli, these visual pathways can still process visual stimuli (Berti & Rizzolatti, 1992). Neuroimaging of these patients have shown that the FFA and the amygdala – an area of the brain that is associated with the processing of danger and emotionally related events – were activated by fearful faces that were not consciously perceived (Vuilleumier et al., 2002).

Taken together, it therefore appears that even when the conscious awareness of a face is disrupted, by either masking face stimuli in healthy participants or by a neuropsychological disorder of vision, the presence of a face is processed to some extent.

1.2.4 Conclusion: are faces processed automatically?

In line with the main criteria proposed by Posner and Snyder (1975a, 1975b) for automatic processing, it does appear that detecting a face is rapid being faster than the detection of other objects, regardless of the number of distractors, indicating that face processing is capacity free. The presence of a task irrelevant distractor face in a task also appears to cause an interference effect on a task, suggesting that face processing may be involuntary. Furthermore, evidence from neuroimaging studies and patients with brain damage indicates that the mere presence of a human face, even when not consciously perceived can activate the FFA. Thus research suggests that faces are processed automatically.

1.3 Selective Attention to Faces:

The previous section focused of the question whether face processing is automatic. A related but different issue is whether faces have a preferential bias to engage the mechanism of selective attention (Palermo & Rhodes, 2007). In the forthcoming section the thesis will examine whether the presence of the human face is more likely to engage and retain selective attention and whether this is done in preference over other objects.

1.3.1 Emotional Faces?

A large body of research has focused on attention capture by emotional faces; consequently, data from various tasks indicate that attention is captured by emotional, predominantly angry facial expressions. One such task is a modification of the typical spatial cueing paradigm (Posner, 1980; Posner & Cohen, 1984). In the classic task, attention is involuntarily or reflexively orientated by the onset of an irrelevant and non-predictive spatial cue to a possible target location. A target is then presented. Typically, when the interval between the cue and onset of the target (stimulus-onset asynchrony; SOA) was less than 300ms, targets at a cued location (valid cue trials) are detected faster than targets at a non-cued location. Conversely, when the cue is non-predictive of a target (invalid cue trials) reaction times are slow.

This methodology has been applied to the study of attention capture by faces. For instance, Fox, Russo, Bowles, and Dutton (2001) demonstrated that in high trait anxiety participants, when an angry face was used as the non-predictive cue to a possible target location on invalid trials, participants took longer to detect a target compared to when either happy or neutral faces were used as the cues. Fox et al. concluded that these results might reflect a delay in the ability to disengage attention from the location of the threat stimuli. Similar results have been observed using the dot-probe task (e.g. Mogg & Bradley, 1999). In a pictorial version of the task, two pictures of faces were presented simultaneously side-by-side, with one face having an emotional expression (e.g. happy or angry) and the other being neutral. After the offset of the faces, two dots were presented at the location of one of the faces; participants were faster to respond to probes that were at the previous location of the angry faces than neutral or happy.

In a flanker task (Eriksen & Eriksen, 1974), where participants made judgements about a target whilst ignoring flanking stimuli, reaction times to a target often vary depending on the nature of the flanking stimuli, targets with compatible flankers were responded to faster than targets with incompatible flankers – the flanker compatibility effect.

Importantly, Fenske and Eastwood (2003) showed that angry faces had much smaller compatibility effects than responses to happy faces, indicating that angry faces were less vulnerable to the effect of peripheral information. In a similar vein, in a pictorial emotional stroop task, naming the colour of an emotional face took significantly longer than naming the colour of a neutral face (Honk, Tuiten, De Haan, van den Hout & Stam, 2001).

Extending this, not only are emotional faces prioritised for rapid detection but words that contain threatening content (Dijksterhuis & Aarts, 2003) and images of snakes and spiders are attended more rapidly than other objects. This apparent predisposition to detect danger within an environment has been labelled as the 'threat-superiority effect' and has been linked to a specialised region within the amygdala that selectively responds to fear (Öhman & Mineka, 2001). From an evolutionary perspective an attentional bias towards threatening stimuli has many important adaptive benefits. For instance, an angry face may signal danger and the ability to reflexively orientate towards a threatening stimulus such as an angry face may be crucial for a person's survival. However, given the importance of the human face for social interaction, the ability to detect a face, regardless of it's facial expression, is one of the most important aspects of face processing (Burton & Bindemann, 2009). In terms of understanding attention capture by faces, if an emotional face is detected faster than another stimulus, it cannot be determined whether attention was captured by the face or by the emotion of depicted on the face. Given this, the face stimuli in the current thesis and remaining literature review will focus on faces that do not contain emotional expressions i.e., neutral faces.

1.3.2 Is there an attentional bias for faces when competing with other objects for attention?

Faces appear to have an advantage in capturing attention when they are in competition with other objects (Palermo & Rhodes, 2007). One paradigm that is used to examine this is the Change Blindness flicker paradigm (CB; Rensink, O'Regan & Clark, 1997). In this paradigm, viewers are often very poor at detecting a change between two images, if they are separated by a transient or brief blank field (imitating an eye saccade). This appears to make the images appear to flicker. Attention is assumed to play a large role in CB because when an object in the scene is pre-cued, a change in that object is more likely to be noticed than other non-cued objects. Using this methodology, Ro et al. (2001) showed that changes in faces were noticed more often and more accurately than changes in other objects. Specifically, they placed six items, each from a different object category (e.g. food, appliance etc.), in an array around a central fixation point. One of these items was always a human face. In each object category, the items where chosen to be maximally different from one another in terms of their shape. Conversely, in the face category, the faces were chosen to be maximally similar to one another (e.g., same face shape and all faces were of the same sex, women). This was done to ensure that any detection advantage for faces could not be attributed to the faces being more different from one another than the items in the other object categories. In the experiment, on half of the trials one of the items in the array would change between the two images. Participants were asked which, if any, of the categories had changed. The results demonstrated that changes in faces were more readily noticed than changes in the other objects. The results from Ro et al. therefore indicate that when in competition with other objects, faces do appear to attract attention more than other objects. Thus, because of this preferential attention to faces, face-related changes were more easily detected compared to non-face stimuli. In a similar study, when one non-face item was presented in an array of faces, changes in the object were noticed faster than changes in faces. This led to the suggestion that the detection advantage observed for faces in Ro et al. was the result of a bias towards attending a novel stimulus presented in an array rather than an attentional bias towards faces (Palermo & Rhodes, 2003). However, it may be that although there is an attentional bias towards faces in the presence of other objects, this bias is overridden when faces are in competition with novel or unique stimuli

Further evidence for an attentional bias for faces when in competition with other stimuli comes from research using patients with visual extinction (see section 1.2.3.). For example, in one study Vuilleumier (2000) presented extinction patients with a series of faces, scrambled faces, meaningless shapes and written names. The stimuli were presented in both unilateral (one stimulus in one side of the visual field) and bilateral (two stimuli: one in the left visual field, one in the right visual field) trials. Participants were then asked to report and locate the stimuli. The results demonstrated that there was no significant difference in unilateral stimuli missed in the left than right hemifield. Additionally, consistent with extinction, the results demonstrated that in bilateral trials, with a non-face stimulus in the ipsilesional visual field of the damaged hemisphere. However, a face presented in the contralesional side of the visual field to the damaged hemisphere was less likely to be extinguished than the other stimuli presented there. This finding was observed across all participants suggesting the unique pattern of effect that faces have compared to the other stimuli used in the experiment. Importantly, these results indicate that faces appear to have a special advantage over other stimuli in visual attention.

Another line of evidence for a face bias when in comparison to other objects is that new-born babies have been shown to prefer to look at face-like stimuli configurations than other visual stimuli. For instance, new-borns will track with their head and eyes slowly moving schematic face stimuli further than they will track scrambled faces or blank stimuli (Goren, Sarty, & Wu, 1975). Using photographs of faces, Cassia, Turati & Simion (2004) demonstrated that new-borns orientated more frequently to upright photographs of faces than they did to inverted and scrambled faces. Taken together, the behavioural and neuropsychological evidence provide strong evidence in favour of faces having an attentional advantage over other visual stimuli in visual processing and their ability to orientate attention.

1.3.3 Do Faces Capture Attention in a Bottom-Up Stimulus-Driven Manner?

Other methodologies have also been applied to the study of attentional preferences for faces. Specifically, the inhibition of return effect (IOR), which was first observed in Posner and Cohen’s (1984) seminal paper when examining the spatial cueing paradigm. In this, they found that when the SOA was greater than 300ms, targets presented at the cued location were detected significantly slower than targets presented at the non-cued location. They referred to this inhibitory tagging effect as IOR and proposed that its purpose is to encourage visual orienting toward novelty, by preventing orienting toward previously attended locations, thus maximizing the experience of the visual world (Posner & Cohen, 1984; Klein, 1988; Klein & MacInnes, 1999). Implicit in IOR is that it would not occur if attention remained at the cued location. To prevent this from happening, studies will often cue, by flashing, the central fixation point after the previously cued location. Additionally, an important condition of the IOR is that the effect only occurs if the allocation of attention to that location was involuntary, or *exogenous*. IOR does not occur following a shift of attention that was directed by an endogenous cue (e.g. an arrow). In their experiment, Theeuwes and Van der Stigchel (2006) applied the notion of the IOR to the question of whether faces capture attention. In their experiment they simultaneously presented two flashed objects (a face and a non-face) on either side of a central fixation point. Following this, participants made speeded eye movements towards a target location as indicated by a central arrow. They observed that there was a delay in response to locations that had been previously cued by a face stimulus but not to locations that had been previously cued with a non-face stimulus. This finding was interpreted in terms of faces giving rise to the IOR. Specifically, in order for the effect to have occurred, attention must have first been involuntarily shifted to the location of the face first, strongly suggesting that the faces had involuntarily captured attention.

Furthermore, as will be discussed in more detail in Chapter 3 (section 3.1), Langton et al. (2008) found that when a face was presented in an array of distracting items, participants' took longer to determine that a non-face target was present, compared to when it was not present. Langton et al. concluded that the presence of a face had captured attention and therefore slowed the visual search for the target. Taken together, the evidence demonstrates that faces are not just preferential in our attention system in comparison to other objects, but that they can automatically capture attention in a bottom-up stimulus-driven manner.

1.3.4 Do Faces retain attention?

Whilst the aforementioned research has successfully demonstrated that faces are preferential when in competition with other objects, and that they capture attention in a bottom-up stimulus-driven manner. Further investigations have shown that faces also retain attention more than other objects. For instance, Bindemann, Burton, Hooge, Jenkins & de Haan (2005a) examined this notion using a go/no-go paradigm. On go trials, a green dot signalled to participants to give a response to a peripheral target. This required participants to disengage and switch attention away from the dot and to the target. Conversely, on no-go trials a red dot signalled participants to give a response that initiated the next trial. However, superimposed behind the central dot could either be an upright face, inverted face, a fruit or a blank background. They proposed that if faces had an advantage in retaining attention, reaction times to trials that had a face superimposed behind the dot would be slower compared to reaction times for trials that had other objects superimposed behind the dot. Their results demonstrated exactly this effect; attention was harder to disengage and shift away from faces than the inverted faces and other objects superimposed behind the dot. Bindemann et al. did suggest that the results could potentially be demonstrating attention capture by faces rather than retention. To eliminate this possibility, they extended the SOA between the central stimulus and target to ensure that attention would be directed to both faces and non-face stimuli. However, the same effect remained, demonstrating upright faces delayed responses to the target more than other object categories. Although this experiment does not demonstrate attention capture, it does fit with previous findings suggesting a relationship between faces and attention (Bindemann et al., 2005a). A more recent study yielded a similar result (Ro et al., 2007), however, this will be discussed further in Chapter 2.

1.3.5 Control of Attention to Faces

Whilst a good deal of research has shown that faces capture attention over other objects, less work has explored the way in which attention is under top-down control in the presence of a face. Using a spatial cueing paradigm, Bindemann et al. (2007) examined the extent to which an observer can control attention in the presence of a face. They presented participants with a face and an object (e.g., image of a tap) concurrently, either side of a fixation point. After varying SOAs the stimuli were removed and a target probe presented at either the face or object location. In line with previous research, they observed that at all SOAs, when the face and object equally predicted target location, responses to targets locations that previously contained a face were significantly faster, indicating a face bias. However, when the face predicted the target on only 25% of the trials there was an object bias instead, indicating that a face bias can be overturned by endogenous control. To further this, in another experiment they examined whether exogenous and endogenous mechanisms were simultaneously working in the task. They repeated the experiment with faces and objects equally predicting target location. However, they also ran two additional experiments, in one of these faces predicted the presence of the target on 75% of all trials, however in the other, objects predicted the presence of the target on 75% of trials. Again, they showed that when an object cue was more predictive, an object advantage was observed. However, they also observed that when faces were more likely to predict the location of the target, responses to the target were significantly faster than when objects were more likely to predict its location. They proposed that this advantage was the result of an additive effect of attention capture by faces and the endogenous mechanisms of attention. Specifically, when the object cue was mostly predictive of the target, the presence of a concurrent face may have also drawn attention slowing down responses to the predictive object cued locations. Taken together, what these results demonstrate is that although attentional biases for faces can be overridden by endogenous control of attention, they cannot be completely eliminated. Rather, endogenous control of attention to objects may be slowed down by exogenous control of attention to faces.

How faces are detected has been the focus of an extensive body of research. As a result, faces appear to be automatically processed. It is also now well established that faces are preferential when in competition with other non-face stimuli, similarly that they rapidly capture and retain a viewer’s attention within a visual field and that although endogenous attention can override attention capture by faces, a face bias will always underpin this effect. As such, many important insights have been obtained from the study of face processing.

1.4 Are Bodies Special Too?

Under natural viewing conditions the human face is rarely perceived in isolation, it is attached to a substantial cue, the human body (Bindemann et al., 2010). Given that faces and bodies co-occur and the qualitative similarities between them as noted earlier, it seems reasonable to suggest that human bodies as visual stimuli may share some of the same attentional processing advantages that faces do (Slaughter et al., 2004). Remarkably, whilst the behavioural evidence is extensive for attention capture by faces, there is comparatively less behavioural evidence of attention capture by bodies and the evidence to-date of a similar advantage for human bodies lags considerably behind. A limited number of research studies have shown behavioural effects for bodies that have been found with faces. In the previous section the question of whether faces are subject to enhanced, attentional processing advantages was addressed using criteria from Posner and Snyder's (1975a, 1975b) main criteria for automatic processing and evidence of selective attention to faces. Given the dearth in the behavioural literature regarding attentional biases for bodies, the following section will systematically examine the behavioural literature in order to address whether bodies have attentional processing advantages, addressing the criteria for automatic processing if possible and where applicable.

1.4.1 Visual Search for Human Bodies?

Although no one study has directly applied the visual search paradigm to address whether bodies pop out, or are detected in parallel with a scene, indirect evidence exists from investigating how attention is distributed during free viewing of natural scenes. Fletcher-Watson, Findlay, Leekham and Benson (2008) presented participants with two natural scenes simultaneously, one of which contained a human body. They examined whether the person would capture attention when it was not the target of search. Participants had their eye-movements tracked whilst free viewing the two images. A clear bias towards person-present scenes was observed, participants spent significantly longer viewing these scenes compared to person-absent scenes. Importantly, the first fixation, which measures items receiving attentional priority in a scene, demonstrated that 27% of all fixations fell on the human body, compared 15% of first fixations that landed on the face. Participants also spent 40% of the time examining the body on the person in the scene. Although not directly comparable to the visual search in an array of items, these results do indicate that when presented surrounding in natural scenes, the human body attracts attention, suggesting that detection might be capacity free (Fletcher-Watson et al., 2008).

However, another study produced a different result (Birmingham, Bischof & Kingstone, 2008). In this study, under a free-viewing task participants eye movements were recorded whilst they viewed different types of natural scenes containing people.

This showed that there was a preference and therefore more first fixations on the eyes with second preference for the face, followed by the body then objects (Birmingham, et al., 2008). This same pattern of data was also observed when participants were told to examine and then describe the scene and also when asked to examine the social content in the scene. These results suggest therefore that the face may have a preference in guiding attention over bodies. Despite this, bodies were the next most fixated on item in the scene after faces, therefore indicating that there may be some attentional preference for bodies, whether this is over faces is hard to discern from these results. However, what these results do demonstrate is that when bodies are presented in scenes that contained surrounding every day natural objects they have a detection advantage (c.f. visual search paradigm arrays), suggesting that bodies may pop out of a scene, and that that their detection is *capacity free.*

1.4.2 Detection Advantages For Bodies: Inattention Paradigms

Similar to faces, it has also been shown that human bodies are generally resistant to the inattentional blindness effect (IB). A primary test of this by Mack and Rock (1998) was made using the traditional IB paradigm. They found that when the critical stimulus was the stick figure of a person, detection was more efficient than schematic critical stimuli: people were noticed 80%, compared to 33% for the depiction of a house and 45% for a schematic tree. In a similar study, Downing, Bray, Rogers and Childs (2004) also found when using the IB paradigm, body silhouettes were more likely to be detected than three other types of controls. Detection of critical body stimuli was also superior compared to silhouettes of human hands, ruling out a basis for the body detection based on the biological nature or familiarity of the critical stimulus. Detection was also greater for bodies than for scrambled versions of the stimuli, therefore ruling out low-level image-based properties, such as local contrast, as an explanation for the detection advantage observed for bodies. Additionally, detection for bodies was superior compared to silhouettes of non-body objects. Interestingly, they did not find an advantage for stick body figures in comparison to scrambled stick body figures. This finding was in contrast to Mack and Rock’s (1998) earlier finding. However, Downing et al. argued that this might have been the result of the stimuli being presented at parafoveal positions in which the visual resolution might not have been sufficient enough to resolve the stick figures. In order to test this prediction, Downing et al. presented the critical stick figure bodies at the fovea, to maximize visual resolution. The results demonstrated that when presented at the fixation point, like the silhouette bodies, the stick figure bodies were also detected more compared to object line drawings. Taken together, these results do appear to suggest that bodies are priority stimuli for attentional selection in comparison to the controls used within the experiments, indicating that their processing may be mandatory and unavoidable.

1.4.3 Do Bodies (and Faces) Engage and Retain Attention More than Other Objects?

More recently, Ro et al. (2007) used a variation of the additional singleton paradigm (ASP; Theeuwes, 1994) to examine search performance and distractor effects for human faces, body parts and other object categories. In in the ASP, which will be discussed in more detail within Chapter 3, attention capture by a stimulus can be assumed if it can be shown that a task-irrelevant distractor is causing a significant delay on target detection times. Rather than using stick figures and silhouettes as previous research had, Ro et al. used photographic stimuli, thus providing a more realistic appraisal of body detection. They presented participants with an array of items (faces, body parts and other objects, such as food) all inside outline colour frames. Participants were required to search the arrays for a green frame amongst blue frames, indicating whether or not the item in the green frame belonged to the object category of a preceding word cue. On half of the trials an item was also presented in a red frame. Using the colour frames allowed them to exploit the concept of a task irrelevant colour singleton capturing attention (Theeuwes, 1994, 1996). More specifically, previous evidence has shown that the search for a target (i.e., here, the green frame) can be disrupted by the presence of a non-target colour singleton (i.e., the red frame), even when it is not relevant for the task (see Chapter 3, section 3.1. for more detail). As such, the use of the colour frames allowed them to examine how faces and body parts are processed after attention had already been allocated to their location (i.e., by the colour frame singleton). Ro et al. found that when a face or body part was inside the red colour singleton frame, participants’ performance on the classification task of non-human objects was significantly slower compared to when the face or body-part was inside the red singleton frame. Thus suggesting that these stimuli could not be ignored, even when they were task-irrelevant. Furthermore, when a face or body-part was presented within the green target frame, classification was significantly faster for these targets than for classification of other objects when inside the green frame. It was concluded that faces and body parts are processed faster and more efficiently than other stimuli. Although, this study did not directly assess whether faces and body parts capture attention it does demonstrate that they are capable of engaging and holding attention compared to other objects (Ro et al., 2007).

1.5 Inversion Effects for both Bodies and Faces

Although the literature regarding attentional biases for the human body is not as substantial as the face literature, the literature that does exist appears to indicate that there may be attentional processing advantages for bodies, similar to those observed for faces. Some of the most consistent evidence of the similar processing of human faces and bodies is that both stimuli show inversion effects (Yin, 1969; Haxby et al., 1999; Reed, Stone, Bozova & Tanaka, 2003). As previously noted, the face inversion effect is the disproportionate drop in reaction times to and recognition of inverted faces, relative to upright faces (Yin, 1969). Undisputed is that inverted faces are harder to recognize than when they are upright because inversion affects holistic processing (Olk & Garay-Vado, 2011). That is, that faces are processed as a whole. Note there is some degree of ambiguity concerning the terms ‘holistic’ and ‘configural’. Many studies use them interchangeably and although we acknowledge there is argument that the terms have slight different connotations, this aim current thesis was not to distinguish between the two. Therefore, for simplicities sake, the term *holistic* shall be used for the remainder of this thesis. Holistic processing is compared to featural processing where features are processed one at a time. As noted previously, face processing is a multi-stage processing beginning with face detection, inclusive of face recognition. Research has shown that face recognition is underpinned by holistic processing because when a face is inverted, the second-order configuration (the variation in different faces in spatial relationships between the features of the face) is disrupted when inverted thus making recognition difficult (Yin, 1969; Tanaka & Farah, 1993). Importantly, holistic processing has also been shown to underpin an earlier stage of face processing, the detection stage (Taubert et al., 2011). Specifically, it has been shown that the detection of inverted faces is disproportionately worse than the detection of upright faces (Taubert et al., 2011; Garrido, Duchaine & Nakayama, 2008; Purcell & Stewart, 1986).

Therefore, holistic processing also operates to extract first-order information from a face, namely the basic attributes of a face (two eyes, above the nose and mouth below the nose) that allow a face to be distinguished from other objects (face detection; Taubert et al., 2011). Therefore, when a face is inverted the first order configuration of the face that makes holistic processing and detection possible is disrupted. As such, it has been argued that holistic processing supports the rapid detection of an upright, but not inverted face in everyday scenes (Taubert et al., 2011). In line with this suggestion, numerous studies have shown that inverted faces do not attract attention (see Olk & Garay-Vado, 2011; Langton et al., 2008; Ro et al., 2003; Devue, Laloyaux, Feyers, Theeuwes & Bredart, 2009; Gilchrist & Proske, 2006). Conversely as demonstrated in this literature review upright faces do appear to capture attention. Importantly, what this shows is that a face inversion effect can be used as a diagnostic for priority processing for faces (as is done so in Experiment 3, Chapter 2).

Altogether, the presence of such an effect is considered to be a hallmark face processing as it is not normally demonstrated for other objects (Carey & Diamond, 1977; Sergent, 1984; Rhodes, Brake & Atkinson, 1993; Farah et al., 1995; Mckone et al., 2001; Murray et al., 2003). Interestingly, a number of studies have examined inversion effects for human bodies (Brandman & Yovel, 2010; McKone et al., 2007; Minnebusch et al., 2009; Reed et al., 2003; Reed, Stone, Grubb & McGoldrick, 2006; Yovel, Pelc & Lubetzky, 2010). Inversion effects for full-bodies have been shown to be as robust and as large as the face inversion effect. However, it has been shown that the body inversion effect appears to be dependent on the presence of the human face to that extent that inversion effects for headless bodies do not appear to exist (Kessler, Walls & Ghuman, 2013; Brandman & Yovel, 2012; Brandman & Yovel, 2010; Yovel et al., 2010; Minnebusch et al., 2009). Nevertheless, as Stein, Sterzer and Peelen (2012) point out, all of these studies examined inversion effects for headless bodies with discrimination tasks, whereby participants are asked to discriminate between two body postures.

Discrimination tasks utilize different cognitive processes than a detection task (de Gelder, Frissen, Barton, & Hadjkhani, 2003). Specifically, discrimination tasks are a higher level process which involves viewers encoding the body posture, holding it in short term memory and then comparing this to the comparison stimuli (Stein et al. 2012). In detection tasks viewers simply have to detect the presence of the stimulus.

Surprisingly, only one study has examined inversion effects for headless bodies in a detection task. Stein et al. used a continuous flash suppression paradigm (CFS; Jiang, Costello & He, 2007), to render headless bodies, body silhouettes, faces and matched controls invisible. More specifically, in a CFS paradigm, a stimulus presented to one eye is removed from visual awareness, hence made invisible, by the flashing of another image in the other eye (Tsuchiya & Koch, 2005). CFS has advantages over the visual search paradigm as stimuli compete with masks eliminating the problem of target and distractor similarity (Duncan & Humphrey, 1989). Therefore, Stein et al. (2012) presented high contrast masks to one eye of the participant, while the test stimulus was gradually faded into the other eye. The fading in of the stimuli was achieved by linearly decreasing the luminance of the stimuli from white to their original luminance. Stein et al. measured the time taken to detect the stimuli that was faded in. They demonstrated that human bodies, regardless of whether they had a head or were silhouette, were detected faster when they were presented in an upright orientation compared to when they were inverted. The importance of these results is that they demonstrate that inversion effects do exist for headless bodies; however, these may be restricted to detection tasks rather than discrimination tasks.

1.6 Biological Basis for Face and Body Processing

In line with the abundance of behavioural evidence indicative of attentional processing advantages for faces and bodies, there is much neuroscientific evidence to suggest that dedicated neural systems, which selective respond to the representation of faces and bodies, underlie these processing advantages. A system within the brain in which distinct neural areas represent the processing of all objects categories appears implausible. However, evidence does suggest that there may exist distinct neural systems that selectively respond to some stimuli (Downing et al., 2004). Therefore, evidence in favour of a biological basis for the processing of faces and bodies provides further evidence of the specialization of these stimuli in the visual system.

1.6.1 Neuroimaging: Faces

The use of functional imaging has facilitated the study of cortical specialization in the human brain, with the advantages of having a high spatial resolution and large sampling area (Kanwisher et al., 1997). Consequently, strong evidence supporting the notion of dedicated architecture for face detection and processing comes from Neuroimaging evidence. Specifically, three distinct cortical regions have been localized in the human visual extrastriate cortex that selectively respond to faces as compared to other object categories (see Figure 4). These regions are the posterior superior temporal sulcus (pSTS; Phillips et al., 1997; Puce et al., 1998), the inferior occipital gyrus (occipital face area (OFA); e.g. Puce et al., 1996; Gauthier, Skudlarski, Gore, & Anderson, 2000), and the most commonly studied, mid-lateral side of the mid-fusiform gyrus (fusiform face area (FFA); Kanwisher et al., 1997). They are generally thought of as the *core* of face perception, representing a system predominantly for the analysis of the human face (Haxby, Hoffmann & Gobbini, 2000; Ishai, 2008; Rossion, 2008).

Figure 4 Schematic representation of the location of the fusiform face area (FFA), occipital face area (OFA) and posterior superior temporal sulcus (pSTS) in the brain. Adapted from Huang et al. (2004) Individual differences in cortical face selectivity predict behavioural performance in face recognition.

The pSTS, although not as well examined as the FFA and OFA, has been shown to have a strong preference for the perception of dynamic facial movement, e.g. eye and mouth movement (Pitcher et al., 2011; Puce et al., 1998), and have suggested functionality in expression processing (Winston, Henson, Fine-Goulden & Dolan, 2004; Andrews & Ewbank, 2004; Engell & Haxby, 2007; Said, Moore, Engell, Todorov & Haxby, 2010).

Research has also demonstrated that the right OFA is a component essential to the face perception network, with a preference for responding to face parts (Pitcher et al., 2011).

For example, in one fMRI study, Liu, Harris and Kanwisher (2010) examined the response of the OFA to the presence or absence of face parts (eyes, nose and mouth) and the response to the basic T-shaped face configuration those parts make up (Liu et al., 2010). They measured responses to face configuration by replacing the face parts with blank ovals. From the results, they concluded that the OFA responded to the presence to real face parts, though it did not respond to the configuration of the parts (Liu et al., 2010). Consistent with this, evidence from Transcranial Magnetic Stimulation (TMS) demonstrated that when repetitive TMS was applied to the right OFA, it selectively disrupted the discrimination of faces parts. However, it did not disrupt the participants' ability to discriminate faces on the basis of spacing between parts (Pitcher, Walsh, Yovel & Duchaine, 2007). In addition, the study showed that only the TMS pulses delivered at early stimulus onset (60 and 100ms) affected the face-part discrimination, indicating that the OFA is involved at an early stage of the face-processing stream (Pitcher et al., 2007). Gauthier et al. (2000) therefore suggested that that the OFA may function in face detection. In line with this suggestion, a recent fMRI study found activation in the OFA when participants were engaged in a face detection task (Nestor, Vettel & Tarr, 2008).

Importantly, the most robust findings for a face-selective module in the human brain have been shown in the lateral side of the mid-fusiform gyrus, named the fusiform face area (FFA; Kanwisher et al., 1997). Early investigations using Positron emission tomography (PET) in 1990's revealed greater activity in the fusiform region during face viewing relative to other visual objects (Sergent, Ohta & MacDonald, 1992), and also during a face-matching task as compared to a dot-location matching task (Haxby et al., 1991). Consistent with the PET evidence, fMRI has demonstrated greater activation in the FFA in response to faces as compared to a range of stimulus categories. For example, Malach et al. (1995) compared the fMRI activation produced by textures with the activation produced by objects, finding that faces activated the FFA compared to the textures, despite the stimuli having been equated for luminance and contrast. Demonstrations of the activation of the FFA responding more strongly to faces have been also been shown in comparison to flowers (McCarthy et al., 1997), houses, hands and everyday objects (Kanwisher et al., 1997).

1.6.2 FFA: Face-Specific Area or Expertise?

Substantial evidence supports the hypothesis that the FFA is involved in face processing. It has therefore been proposed that the FFA is a domain specific module in the brain for the selective processing of human faces and not other object classes. This apparent neural specificity has been labelled the '*face-specificity hypothesis'* (Kanwisher et al., 1997). However, just because the strongest activation in response to faces are found in the FFA, does not necessarily imply that this is a region dedicated only to face processing (Peelen & Downing, 2005). Indeed, there are several other alternatives interpretations (Kanwisher et al., 1997). First, given the large difference in the low-level image-based properties of face and other stimuli, it is possible that the FFA may be responding to these rather than the faces themselves (Kanwisher et al., 1997). However, evidence from using bistable stimuli shows that the FFA responded more strongly when the face is being perceived than when the vase is perceived (Andrews, Schluppeck, Homfray, Matthews & Blackmore, 2002). Given this evidence, it is difficult to account for the apparent selectivity in the FFA for faces in terms of their low-level visual features (Kanwisher & Yovel, 2006).

Second, the FFA instead may represent an area of fine level discrimination, which arises from extensive experience in individual level discrimination of objects that a viewer has had substantial visual expertise of (expertise hypothesis; Diamond & Carey, 1986). Therefore, the heightened selectivity for faces in the FFA can be explained as a consequence of learning through extensive exposure to faces (e.g., Carey, 1992; Diamond & Carey, 1986; Gauthier & Tarr, 1997; Meadows, 1974). In addition, the expertise hypothesis predicts that the greater visual experience with faces in an upright orientation leads to the behavioural face inversion effect (Diamond & Carey, 1986; Gauthier, Tarr, Anderson, Skudlarski & Gore, 1999; Gauthier & Tarr, 1997). The expertise hypothesis also predicts that FFA should respond more to objects of expertise than other non-expert controls. Object expertise can be defined as the ability to rapidly and accurately recognize objects at a subordinate level (Tanaka & Gauthier, 1997). Evidence in favour of the expertise hypothesis comes from converging lines of both behavioural and neuroimaging research. Specifically, Gauthier and Tarr (1997) examined changes in sensitivity with novice and expert participants to configural transformations with "Greebles" - novel animal-like stimuli. Expert participants were given extensive training to become experts in Greeble recognition. Their results demonstrated that the experts were significantly faster at detecting subtle changes to the configuration of the parts of the Greebles compared to the novices. Additionally, this effect was not demonstrated with the experts when the Greebles were inverted. Taken together these results suggest that with sensitivity to the configuration of objects can be learnt throughout extensive experience with that stimulus. Furthering these findings, a number of studies also demonstrated the inversion effect, which was once thought to be unique to faces, for other objects of expertise. Specifically, Diamond and Carey (1986) found that expert dog handlers when shown pictures of dogs demonstrate the inversion effect, however, non-expert dog handlers do not show the same effect. In a similar study, Curby, Glazek and Gauthier (2009) presented participants with a range of experience in identifying cars with upright and inverted faces and cars. They divided the participants on the basis of the car expertise using self-report measures. The results demonstrated that for both experts and novices there was an inversion effect for faces, whereas, only the car experts demonstrated an inversion effect for cars. Taken together, the presence of an inversion effect for non-face objects provides behavioural evidence in favour the FFA being an area within the brain for the individuating between homogenous objects that a viewer has had substantial experience and viewing of.

Evidence from fMRI research has produced conflicting results in attempts to establish evidence for the expertise hypothesis. Gauthier et al. (1999) gave participants extensive visual training with Greebles. Both these experts and novices were then scanned using fMRI during passive viewing of the greebles. In the experts, they observed that there was more activation in the FFA for upright greebles relative to inverted greebles. Moreover, the FFA was also activated more in the experts than the novices during passive viewing of the greebles (Gauthier et al., 1999). In another study, Gauthier et al. (2000) examined whether long-term expertise with birds and cars would activate face selective areas in the brain. They measured responses in the FFA and OFA whilst car and bird experts performed location and identity matching tasks. They observed that there was more activation in the FFA and OFA when the experts were performing tasks with stimuli from their area of expertise versus control objects. In addition to this, their participants also performed a behavioural task outside of the scanner whereby participants judged whether two birds were from the same species or cars from the same model. Interestingly, Gauthier et al. (2000) state that their most remarkable result was a strong significant correlation between the behavioural task and activation in the FFA when during performing on the location-matching task. However, as Kanwisher and Yovel (2006) acknowledge, this result was not observed with the identity-matching task (which involved discriminating between participants stimuli of expertise) where the expertise hypothesis would predict it.

Instead, as Xu (2005) also point out, Gauthier et al. (1999, 2000) both used a blocked fMRI design, whereby in each block every trial belonged to the same category (e.g., all cars or birds). In this, participants can anticipate the next trial and an expert might find a stimulus from their area of expertise more interesting to look at than a stimulus outside of their area. Previous research has shown that other objects can activate the FFA, although less than faces (Kanwisher, et al., 1997; Haxby, Gobbini, Furey, Ishai, Schouten & Pietrini, 2001). Therefore, as Xu states, even if the FFA were not sensitive to objects of expertise, varying levels of attention to objects of expertise would result in a higher FFA response to these stimuli than non-expertise objects. Therefore, these results may instead reflect an increased amount of attentional engagement towards the experts' objects of expertise. As Xu (2005) also points out, there is another specific drawback from the aforementioned studies. All used stimuli that resembled or contained face-like parts. Specifically, Gauthier et al. (1999) and Gauthier and Tarr (1997) used greebles, which are structurally similar to, faces. Gauthier et al. (2000) used birds, which have eyes and mouths similar to that of the human face. Gauthier et al (2000) and Curby et al (2009) used cars, which have headlights that may resemble symmetrical eyes of a face. Accordingly, these stimuli represent a poor choice in order to distinguish between the face-specificity and the expertise hypothesis (Kanwisher & Yovel, 2009). As such, the validity of the expertise effect remains equivocal and open to debate.

Indeed, contrasting the expertise hypothesis are several studies that have shown no significant increase in activation in the FFA for objects of expertise. For example, Grill-Spector and Malach (2004) presented briefly masked faces and a range of stimulus exemplars, including houses, guitars and flowers to non-expert participants. They examined the question of whether the FFA is purposed for the detection and identification of faces. Using both behavioural and fMRI measures they compared responses on trials where viewers succeeded versus trials on which they failed to detect and identify face stimuli. They reasoned that higher response in the FFA during successful detection of face trials as compared to face detection miss trials would indicate the FFA in face detection. The same logic was applied to face identification. They also examined whether the FFA is engaged in the individuation of objects that viewers that had extensive visual expertise of (car experts). Their results demonstrated that there was a significantly greater response in the FFA when participants detected a face compared to when they did not. There was also a stronger response when identifying the face compared to when it was just detected. They argued that these results suggest the involvement of the FFA in both detection and identification of faces. Importantly, their data provides evidence against the expertise hypothesis. First, they observed no correlation between the response in the FFA and the identification of non-face objects. Second, given that the response in the FFA was correlated with both detection and identification of faces they suggest that the FFA is not exclusively engaged in the individuation between faces. Third, they observed no correlation between the viewer's expertise and FFA response. Their data therefore provides no support for the hypothesis that the FFA is concerned with individuating between objects of expertise. Subsequently, several studies have also reported non-significant results for expertise specific stimuli activation in the FFA (Moore, Cohen, & Ranganath et al., 2006; Yue, Tjan, & Biederman, 2006; Op de Beeck, Baker, DiCarlo, & Kanwisher, 2006).

1.6.3 Neuropsychological Evidence

In line with the abundance of neuroimaging evidence, evidence from patients with brain damage also supports the notion of a face selective brain region. Indeed evidence from experiments showing a double dissociation between face and objects recognition suggests that some brain mechanisms serve object recognition better than face recognition and vice versa. For example, patients with prosopagnosia are generally impaired in facial recognition, but have the ability to recognize other objects. This disorder normally manifests after acute brain damage with the lesions normally located in the FFA (Damasio, Damasio, & Van Hoesen, 1982; Sergent & Signonet, 1992; Barton, Press, Keenan, & O'Connor, 2002). Exemplifying this, Farah, Levinson and Klein (1995) had normal control participants and patient L.H. perform a recognition memory task with faces and other controls (e.g., ruler, television). Participants viewed a randomly shuffled deck of study cards each with a stimulus on it. They were allowed to study each card for 5 seconds (LH was allowed 10 seconds). The viewed cards then formed part of a larger set of cards of which the participants were shown in a random order. They indicated whether the card they were viewing was either ‘old’ or ‘new’. They observed that the normal controls performed the same with the face and non-face objects, whereas LH performed worse on the face than non-face objects. On the other hand, one of the most remarkable sources of this evidence is a very unusual agnostic Patient, CK, who after a car accident could recognize faces just as well as controls, but who was severely impaired at object recognition (Moscovitch et al., 1997). For example, when CK was shown paintings from the artist Arcimboldo, who created pictures of faces made up of objects, it was reported that CK was able to recognize the faces, but not the objects. Although this evidence is based on face and object *recognition,* not detection, it does suggest evidence in favour of the existence of specialized neural modules for face perception.

In Sum, neuropsychological lesion studies have shown that face recognition can be impaired independently from object recognition (prosopagnosia) and vice versa (object agnosia). This evidence alone strongly suggests functional divisions that support a face-specificity hypothesis (McKone et al. 2007). Supporting this, evidence from experimental training studies, and long-term expertise studies, appear to suggest that expertise does not mediate activation in the FFA. Indeed, converging evidence appears to suggest that the neural mechanisms that are engaged during face perception are distinct from those engaged in object perception, strongly indicative of a dedicated architecture for face processing and the face-specificity hypothesis (Kanwisher & Yovel, 2006).

1.6.4 Bodies: Neuroscientific Evidence

Whilst there is strong evidence for neural specificity for face like objects, there is also mounting evidence that the human body may also have a neuronal architecture in the brain that is highly selective to representations of the human body. Data from recent studies of the neuronal representation of human body stimuli have converged around two strongly selective areas of the brain that are responsive to visual representations of the human body. Those areas are the extrastriate body area (EBA: Downing, Jiang Shuman & Kanwisher, 2001), located on the posterior end of the inferior temporal sulcus, adjacent to the OFA, and the fusiform body area (FBA; Peelen & Downing, 2005; Schwarzlose, Baker, & Kanwisher, 2005), located on the lateral fusiform gyrus, where it overlaps with the FFA (see Figure 5). The discovery of these regions has led to an enormous amount of amount of subsequent research (Downing & Peelen, 2011).

Figure 5 A schematic representation of the location of the extrastriate body area (EBA) and fusiform body area (FBA) in the brain. From Taylor, Wiggett and Downing (2007) Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas.

Downing et al. (2001) provided the first evidence from fMRI for the body-selective EBA over a series of experiments. In these experiments viewers were scanned whilst viewing objects from different categories, human bodies and body parts. They observed significantly stronger responses in the right lateral occipitotemporal cortex when subjects viewed human bodies and body parts compared to when the objects were viewed. Interestingly, it has been proposed that the EBA is involved in body detection rather than other body processes, such as identification (Chan, Peelen & Simion, 2004; Hodzic, Kaas, Muckli, Stirn & Singer, 2009). For example, von Koningsbruggen, Peelen and Downing (2013) used TMS to examine the role of the EBA in the detection of human bodies in natural scenes. They delivered TMS and sham stimulation to the right EBA of participants during viewing of natural scenes that contained cars or people. Participants indicated then whether the scene they had viewed contained a target (car or person depending on the block). Their results showed that stimulation to the right EBA reduced the participants' ability to detect the human bodies in the scenes compared to the sham stimulation and compared to the car targets. They concluded from their results that the right EBA selectively contributes to the detection of human bodies in natural scenes.

The above findings also support earlier findings of another body selective region of the brain. For instance, Kanwisher et al. (1997) first observed that responses in the FFA were greater to bodies than they were to other objects; however, this was still significantly below the responses to faces. In another study fMRI study, Bartels and Zeki (2004) measured participants' brain activity while they free viewed a trailer for a James Bond movie. They also had different observers view the same clip four times, rating it each time on the intensity of four attributes (colour, faces, bodies, and language). A correlation between the intensity of perception of colour, faces, bodies and language and brain activity revealed that the high intensity of a feature (colour, faces, bodies, and language) correlated with activity in the area specialized to process it. Specifically, the perception of a scene with a rating of high face intensity correlated with activity in the FFA. Further, the perception of scenes with a rating of high human body intensity correlated with activity in two regions, the EBA, but also in another area overlapping with the FFA, suggesting the presence of another selective body region (Bartels & Zeki, 2004). However, in the high intensity body scenes, the stimuli were not controlled and faces were still visible, as a result it was unclear whether this reflected response to faces or bodies (Peelen & Downing, 2005).

The first rigorous evidence for this body selective region (FBA) came from using fMRI and analyses of individually defined regions-of-interest, during free viewing of photographs of faces, bodies, tools and scenes (Peelen & Downing, 2005). It was found that the two areas of the fusiform gyrus that were defined for face- and body- selectivity showed differential patterns of selectivity. The face-defined area (FFA) was found to respond more to faces than to bodies and other stimuli. Similarly, the area defined for body-selectivity (subsequently named the FBA) showed more activation in response to bodies than to face and other stimuli. As such, their results demonstrated that two activations occupy close but distinct locations of the mid-fusiform gyrus (Peelen & Downing, 2005). Subsequently, numerous fMRI studies have consistently demonstrated that the EBA and the FBA selectively respond to both schematic and photograph images of bodies or body parts significantly more than other objects, object parts and many other control stimuli (Chan et al., 2006; Downing et al., 2001; Peelen

& Downing, 2005; Schwarzlose, Baker & Kanwisher., 2005; Spiridon, Fischl, & Kanwisher, 2006; Weiner & Grill-Spector, 2010).

Interestingly, evidence from neuroimaging suggests that the cortical systems that process bodies and faces might be similar (Taylor, Wiggett & Downing, 2007; Peelen & Downing 2005; Spiridon, et al., 2006; Cox, Meyers & Sinha, 2004). Specifically, a functional analogy has also been drawn between the FBA and FFA, indicating that the FBA and FFA have a general preference for whole body or whole face processing, respectively (Taylor et al., 2007). Furthermore, a functional homogony has also been drawn between the EBA and OFA (Taylor et al., 2007), with both areas preferentially responding to parts of the body or face, respectively (Liu et al., 2010, Nichols, Betts & Wilson, 2010; Pitcher, et al., 2007; Taylor et al., 2007).

Altogether, evidence from neuroimaging and behavioural studies strongly suggests that face and body processing involves different processes to that of object processing.

1.7 Faces and Bodies in Natural Backgrounds

This review has provided a comprehensive account of the literature regarding attention and what is currently known about attention capture by human faces and bodies. Owing to the social and biological significance and neural specificity of the face and body, much research has examined whether these stimuli have attentional processing advantages and whether they are able to automatically capture a viewer's attention. However, whilst the reviewed evidence in favour of attention capture by faces is extensive, there is comparatively less behavioural evidence in favour of a similar attentional advantage for human bodies. Notably, nearly all of the aforementioned behavioural literature and much of the neuroscientific literature on face and body studies have been criticized for using artificial stimulus displays, such as simple uniform colour backgrounds to present stimuli (Bindemann et al., 2010). However, under real-world circumstances faces and bodies are not perceived in uniform colour backgrounds, they are normally surrounded by natural or domestic objects (Lewis & Edmonds, 2005). As such, previous findings of face and body detection advantages are confined to these impoverished displays, hence there is relatively little evidence concerning how both faces and bodies are detected and processed in the real world (Bindemann et al., 2010). It is possible that when presented under more realistic circumstances, e.g., in natural backgrounds, advantages for these stimuli may no longer be present. Therefore, in order to fully understand face and body processing it requires that these processes be studied in an ecologically valid context, such as complex natural scenes.

In line with this, one study has directly examined how the human face, body and full-body were detected when presented in natural scenes (Bindemann et al., 2010). In their seminal paper, Bindemann et al. (2010) provided the first direct test of the independent contribution of the face and body to the ability to detect a person in natural scenes. For this purpose, they presented participants with natural scenes and measured the time taken to search the scenes for a target person or part of a person. Some of the scenes presented the full-body of a person (face and body), whilst other scenes showed only a face or body in isolation. They used natural scenes to present the target stimuli, for example, using pictures of living rooms, inside and outside of university buildings. They reasoned that if faces facilitate the detection of a person in a natural scene, then the detection of the face should be fastest and unaffected by whether the body was also present or not. In a similar vein, if bodies drive the mechanisms that lead to person detection in natural scenes then detection should be fastest when the body is presented independently of the face. However, they also predicted that if both faces and bodies contribute to detection, then full-bodies, when the face and body are combined, should be detected the fastest. Their participants searched the natural scenes making speeded target present and target absent judgments based on whether they detected a full-body or part of a person (face or body) in the scene. They measured the speed and accuracy in which participants detected the targets. Their participants' eye movements were also recorded throughout the experiment.

Their results demonstrated that the detection of the body targets was as fast as the detection of the face targets, and that full-body targets (face and body) were detected significantly faster than the face or the body in isolation. The eye movement data added further insight into how the face and body are detected. They observed that although the body received fewer first fixations than the face, overall the body received more fixations than the face, regardless of whether presented together as a full-body or as separate isolated targets.

This led Bindemann et al. (2010) to the suggestion that in natural scenes, the face and body contribute equally to the detection of a person. Therefore, given that human faces have consistently been shown to have attentional processing advantages, this finding could suggest that the body also has attentional processing advantages.

Nevertheless, as Bindemann et al. point out, the larger size of the body targets relative to the smaller face targets may have enabled the bodies to be detected as fast as faces. However, they dismissed this possibility given that bodies were detected as fast as faces, suggesting therefore that size of the targets could not be linearly related to detectability (Bindemann et al., 2010). Target detection was fastest for the full-body targets suggesting that the combination of the face and body may lead to an integration of their respective processing speeds to the extent that the full-body is subject to extra rapid detection and processing.

A closer inspection of Bindemann et al.'s paper reveals a main issue that is worth exploring further. In particular, Bindemann et al. did not directly compare detection rates for faces, bodies or full-bodies against to the detection rates of any other control stimuli. Given this lack of a baseline comparison for the human targets, their results cannot address whether faces, bodies and full-bodies would be detected and processed faster than other stimuli in natural scenes. In summary, the conclusions from Bindemann et al. appear to suggest that both the human face and body are detected rapidly in natural scenes, which may suggest that both stimuli may have attentional processing advantages in natural scenes. However, it is not clear whether this advantage will hold true when in comparison to other stimuli.

1.8 Rationale and Principle Aims of the Thesis/Structure of this Thesis

This section of the review offers an outline of the research in this thesis. As has been discussed, whilst many important insights have been obtained from the study of face and body processing, nearly all of the literature discussed in the above review has used simple artificial displays to present stimuli in. Consequently, any findings of advantages for faces and bodies observed in these simple backgrounds are confined to these simple displays and are therefore limited in what they can reveal about attention capture by faces and bodies in the real world (Bindemann et al., 2010). It is possible that and when presented under more realistic circumstances, e.g., in natural backgrounds, advantages for these stimuli may no longer be present. Consequently, the aim of this thesis was to investigate whether human faces and bodies have attentional priority processing advantages when presented in natural scenes.

The first experimental chapter (Chapter 2) aimed to replicate in part the original experimental finding from Bindemann et al. (2010) of the equal detection rate of faces and bodies when presented in natural scenes using a visual search paradigm. Additionally, these detection rates were also compared to the detection of non –face and –body targets to examine whether detection advantages will be held over other stimuli (Experiments 1 – 3). If human faces and bodies have attentional processing advantages over other stimuli, it would be expected that they would be detected faster and more accurately than other objects in natural scenes. The purpose of Chapter 3 was to provide a stricter test of attention. To assess this, the additional singleton paradigm (e.g. Theeuwes, 1994) was used in which participants had to detect a non-body or non-face whilst ignoring distractor faces (Experiment 4) and distractor bodies (Experiments 5 & 6). Attention capture by faces and bodies, or the face/body distractor effect, was inferred if it was shown that the presence of the distractor had caused a significant delay on target detection relative to when absent. Therefore, if faces and bodies capture attention in natural scenes it is expected that their presence in the scenes should significantly slow the detection times of the target. The additional singleton paradigm was used again in Chapter 4. Two experiments compared the body distractor effect to the distractor effect of control stimuli (Experiment 7 & 8). If human bodies have an attentional processing advantage over other objects then in Chapter 4 it would be expected that the presence of a distractor body should slow detection times to a target more than the presence of control distractors. The final experimental chapter (Chapter 5) investigated the visual properties (e.g. size of target and presence of the face) that may be responsible for eliciting the body distractor effect. It would be expected that if faces capture attention, then the presence of a face in a full-human body should slow responses to the target.

Chapter Two

2 Visual Processing of Faces, Bodies and Full-Bodies Using the Visual Search Paradigm.

2.1 Experiment 1

2.1.1 Introduction

There are recent examples in the literature that suggest that, like faces, bodies may also have attentional processing advantages when presented in natural scenes. Specifically, Bindemann et al. (2010) demonstrated that when presented in natural scenes, human bodies (without the face) are detected as fast as human faces (without the body). As stated in Chapter 1, given the abundance of literature suggesting an attentional bias for faces in visual scenes, this finding may suggest that bodies also have an attentional bias that allows them to be rapidly detected in scenes. Therefore, the key aim of Experiment 1 was to replicate Bindemann et al.'s (2010) finding of the equal detection times of human face and body targets presented in natural scenes.

Importantly, however, Bindemann et al. (2010) did not compare the detection rates of faces and bodies to control stimuli. Therefore, as it stands, it is currently unclear whether a preference for face and body targets will be observed when in comparison with other objects in natural scenes. As such, in addition to replicating Bindemann et al.'s finding of equal detection times for faces and bodies, Experiment 1 also aimed to extend their findings by comparing the detection rates of human face (head, no body attached), body (body, no face attached) and full-body (head and body attached) stimuli against the detection rates for analogous stimuli. Grandfather clocks were chosen as comparable (analogous) stimuli, as they share some similarities with the human form: they have both a face (with regions of high contrast), body, and are of a similar statue. In line with Bindemann et al. and given that nearly all previous research has used artificial displays to present face and body stimuli, all stimuli were presented in natural backgrounds. It is both useful and important to investigate how human faces, bodies and full-bodies are detected under conditions that are a closer approximation of real life in order to better understand how these stimuli are detected and processed in the real world.

Therefore, in Experiment 1, participants were required to search natural scenes for human (human: face-only, body-only, full-body) and grandfather clock targets (clock: face-only, body-only, full-body). Participants completed six blocks of trials: human face-only, human body-only, human full-body, clock face-only, clock body-only, and clock full-body. In each block, half of the trials contained a target (target present scenes), whereas half of the trials did not contain a target (target absent scenes).

To explore Bindemann et al.'s (2010) finding of equal detection for face-only and body-only targets, reaction times (RTs) for human face-only targets were compared to RTs for human body-only targets. If human bodies have processing advantages that are similar to faces, then it would be expected that human body-only targets would be detected as fast as face-only targets. Likewise, the combination of the face and body as a full-body target was compared to the face-only and body-only targets. It was reasoned that, if faces and bodies have processing advantages as independent targets, then when combined as full-body targets, detection should be faster than detection of the independent face-only and body-only targets. The aim of replicating Bindemann et al.'s (2010) original finding was to examine whether body-only targets would be detected as fast as face-only targets. This finding would further suggest attentional advantages for human body-only targets.

Additionally, to determine whether an advantage for human targets is found when these stimuli were compared directly to other analogous stimuli, RTs for the human face-only, body-only, and full-body targets were compared to RTs to the grandfather clock targets. It was reasoned that if faces, bodies and full-bodies have processing advantages over other stimuli, then RTs to detect human targets should be faster than RTs to detect grandfather targets.

2.1.2 Method

2.1.2.1 Participants:

A convenience sample of 42 participants (35F, 7M) between 18-23 years was taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 30 minutes.

2.1.2.2 Apparatus and Stimuli:

Experiment 1 was conducted using a Macintosh MacBook pro 2.4 GHz laptop with a 15inch monitor. Psychopy (1.75.01; Peirce, 2007) was used to present the stimuli and to record reaction times and accuracy. Responses were entered via a response box (Cedrus, Model RB-730). The backgrounds consisted of 24-bit RGB photographs of 30 different scenes (15 inside and 15 outside). These photographs were taken at various locations including the inside and outside Nottingham Trent University, houses and on streets, using a SAMSUNG ES17 digital camera. The scenes measured 1,074 (W) by 768 (H) pixels at a screen resolution of 88.08 pixels/inch, giving them a visual angle of 26.69° × 19.22°, at an approximate viewing distance of 60cm. The target human stimuli were created from photographic pictures of twelve white adults (6 females, 6 males) recruited by the experimenter. All of the human images were cropped out of their original photograph background. Removing the heads from the bodies created the body-only targets and removing body-parts from the neck down created the face-only targets. To produce the target grandfather clock stimuli, twelve photographs of different grandfather clocks were obtained from an open source on the internet and edited in a similar way to the human targets. The face of the clock was defined as the face-only targets (see Figure 6 for an example of human and clock stimuli). All manipulations were completed in ADOBE Photoshop (CS 8).

For the scenes in which a target was present (target present scenes), a face was shown in isolation (face-only condition), a body was shown in isolation (body-only condition), or both the face and body were combined as the full body (full-body condition). Examples of stimuli and experimental conditions are shown in Figure 6. Target absent trials also were included as a control; these scenes contained only the background and no target. These were included to eliminate the predictability in the task demand load (i.e., target present/absent) of the experiment. These trials were not included within the analysis.

In all visual scenes, Psychopy randomly selected a background so that each background was used an equal number of times. In target present scenes, Psychopy also randomly selected a target and its location in the scene. Targets were displayed an equal number of times. As indicated in Figure 6, to avoid any biases towards the same spatial location, targets were presented equally within each vertical third of the screen (left, middle, right) over the course of the experiment.

Human full-body

Human body-only

Human face-only

Clock full-body

Clock body-only

Clock face-only

Figure 6. An illustration of the backgrounds and target stimuli for each of the experimental conditions for Experiment 1.

2.1.2.3 Design

A 2 × 3 within-subjects design was employed with two factors: target type (human vs. clock) and target part (face-only, body-only, full-body). The dependent variables were the reaction time (milliseconds) from the stimulus onset and participant accuracy (i.e., correct target present/absent response). The data were analysed using a 2 × 3 Analysis of Variance. Post hoc analysis (where appropriate) used paired *t* tests with a false-discovery rate correction for multiple comparisons (Benjamini and Hochberg, 1995).

2.1.2.4 Procedure

The experiment was run in six blocks. Each participant completed each block, these blocks were: human-face, human-body, human full-body, clock-face, clock-body, and full clock-body. The target of the search was always dependent on block. The ordering of the blocks was randomized using a balanced Latin-squares design. The trials in all blocks were completely randomized for each participant. Each block was preceded by 8 practice trials followed by 72 experimental trials (36 target-present trials, 36 target-absent trials), totalling 432 trials per participant. Participants were asked to quickly and accurately determine whether or not a target was present in a scene and to make their responses by pressing the corresponding keys on the button box. A trial sequence is illustrated in Figure 7. In each block of the experiment, every trial began with the onset of a fixation cross which was displayed for 1000 milliseconds (ms). The location of the fixation cross was randomly rotated around four possible onscreen locations corresponding to the four corners of the screen. Immediately after this, backgrounds with or without the target were displayed until a response had been made. If no response was made within 3,000ms, the scene was removed and the next trial was initiated.

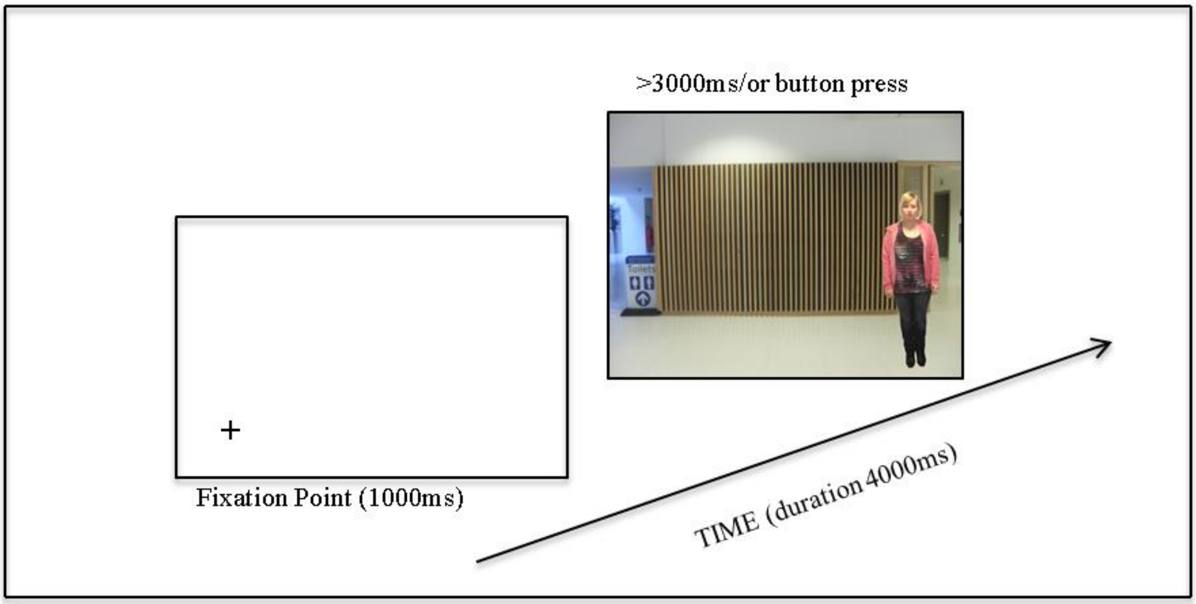


Figure 7. Illustration of the procedure for Experiment 1.

2.1.3 Results

Incorrect button presses and trials where no response was given (none responses) were counted as errors. All errors were removed before analysis of the RT data and included in the accuracy analysis. The median correct target present RTs for each participant in each condition was computed. The average median RTs are displayed in Figure 6, along with the mean errors for each condition.

The error rates were low, with 2.07% of errors made on the target absent trials, 3.9% of the errors made on human present trials and 3.7% of the errors on the clock present trials. As can be seen from Figure 8, error rates were generally consistent with the RTs, with more errors in conditions that participants' performed more slowly (i.e. face-only targets), providing no evidence of a trade-off between speed and accuracy. Accuracy analyses are therefore not reported any further.

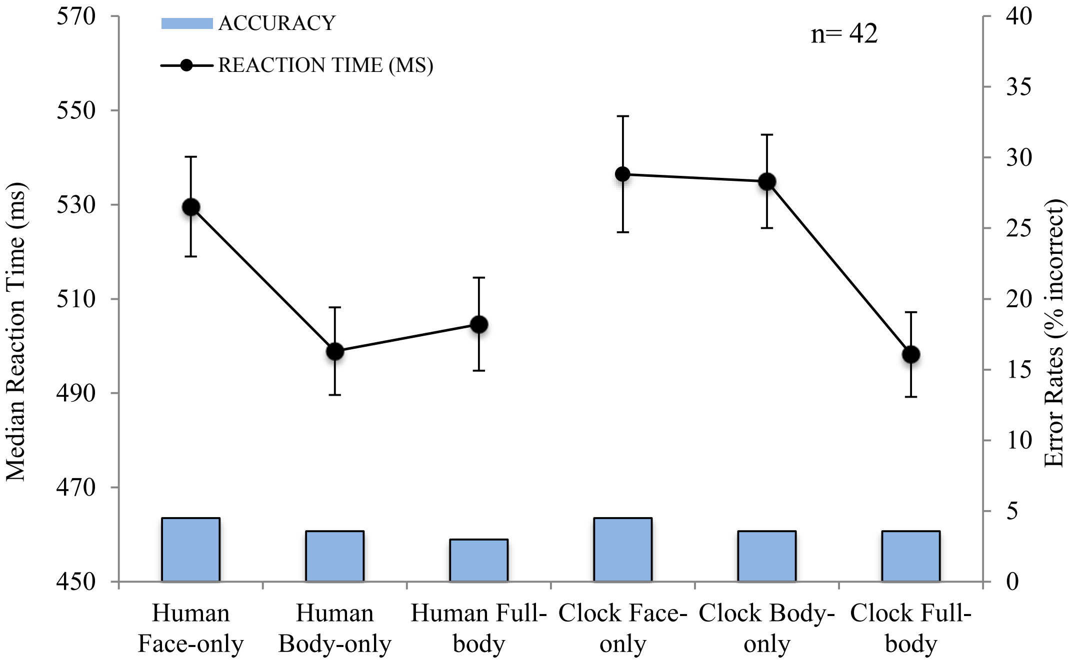


Figure 8. Means of the median correct detection rates and error rates for each condition of Experiment 1 for the target-present conditions. The vertical error bars represent 95% within subject confidence intervals (Cousineau, 2005; Morey, 2008).

Preliminary analyses reveal no main effect or interactions of block order ($p > 0.05$), so it was not considered further. The data was therefore collapsed across all blocks. A 2 × 3 within-subjects ANOVA with factors target type (human, clock) and part (face, body, fully-body) was a conducted on the median RT data. This analysis yielded a marginally significant main effect of target type, $F(1, 41) = 3.93$, $MSE = 9319.67$, $p < 0.05$, $\eta^2 = 0.03$, with faster detection rates for human targets (M=511.04) than for clock targets (M=523.21). The main effect of target part was also significant, $F(2, 82) = 9.97$, $MSE = 20981.53$, $p < 0.01$, $\eta^2 = 0.12$. A significant interaction was also found between target type × part, $F(2, 90) = 9.20$, $MSE = 9920.85$, $p < 0.01$, $\eta^2 = 0.06$.

Paired comparisons using t-tests with the Benjamini-Hochberg α-error adjustment revealed that human body-only targets ($M = 498.90$) were detected significantly faster than human face-only targets ($M = 529.60$; $p < 0.05$, $d = 0.46$). As can be seen from Figure 8, human full-body targets (M=504.63) were detected faster than human face-only targets ($M = 529.60$), however this difference failed to reach significance ($p > 0.05$, $d = 0.41$). There was no significant difference observed between the detection of body-only ($M = 498.90$) and full-body targets ($M = 504.63$, $p > 0.05$).

For the clock target parts, clock face-only targets (*M* = 536.45) were detected significantly slower than clock full-body targets (*M* = 498.20; $p < 0.05$, $d = 0.54$), and clock body-only targets (*M* = 534.96) were detected significantly slower than clock full-body targets ($p < 0.05$, $d = 0.64$). There was no significant difference in detection between clock face-only (*M* = 536.45) and clock body-only targets (*M* = 534.96, $p > 0.05$, $d = 0.03$).

Human body-only targets (*M* = 498.90) were detected significantly faster than both the clock face-only (*M*=536.45, $d = 0.54$) and clock body-only (*M*=534.96, $d = 0.60$) targets. In turn, full-body (*M*=504.63) targets were also detected significantly faster than both the clock face-only (*M*=536.45, $p < 0.05$, $d = 0.50$) and clock body-only (*M*=534.96, $p < 0.05$, $d = 0.44$) targets. No other comparisons were significant (all $p > 0.05$).

2.1.4 Discussion

The results of Experiment 1 are inconsistent with those of Bindemann et al. (2010), who found equivalent detection times for face- and body-only targets, and an advantage for full-body targets. Instead, the results here show that there were detection advantages for human body-only targets over human face-only targets, and although not significant, numerically, human full-bodies were detected faster than human faces. In addition, body-only targets were detected as fast as full-body targets. For the clock target parts, the results show that full-body targets were detected faster than both the body-only and face-only targets. Moreover, the detection of clock body-only and face-only targets was equivalent. Taken together, with the exception of the equal detection times for clock face- and body-only targets (as further discussed below), the general pattern of results is consistent with the larger targets (body-only and full-body) having detection advantages over the smaller targets (face-only). Although target size did not appear to be an issue in Bindemann et al.'s research, they did acknowledge that it could be a potential characteristic that enables faster detection. In line with this, previous research has demonstrated that physically larger objects in a scene capture attention in a bottom-up manner, independent of top-down influences (Proulx & Egeth, 2008; Proulx, 2010). Therefore, the most likely explanation for the detection advantages of the larger targets in Experiment 1 is that target size is related to target detectability and that they are detected faster because they are larger in size. Specifically, at the same viewing distance the full-body and body-only targets were much larger in a scene than face-only targets.

A further finding was that a number of the human target parts were detected faster than the clock target parts. Specifically, both the human full-body and body-only targets were detected faster than the clock face-only targets. Whilst this might suggest that these targets have attentional advantages over clock face-only targets, given that the human targets are larger in size than clock face-only targets it seems more likely that they are detected faster because of their larger size. Thus contributing to cumulative evidence that target size is related to target detectability.

In addition, both the human body-only and full-body target parts were detected faster than the clock body-only target parts. These targets are all of a similar size and it could therefore be argued that the human target parts have detection and processing advantage over the clock bodies. However, considering that it is generally accepted that human faces have advantages over other stimuli (Ro., 2001; Ro, et al., 2007; Bindemann et al., 2005; Langton et al., 2008), then advantages for human face only targets over the clock face-only targets, which are of a similar size, would also have been expected to have been seen. This was not the case. For this reason, it seems unlikely that the faster detection times for the human body-only and full-body targets over clock body-only targets reflects an attentional processing advantage for human bodies and full-bodies. Rather, an alternative hypothesis is that this effect instead represents a disadvantage in the detection of clock body-only targets. In line with this, there was no observed difference between the detection of the clock face-only and body-only targets. Given the evidence that target size appears to be related to target detection, the larger clock-body target being detected as fast as the smaller face target could imply that participants had some difficultly in the detection of these clock body-only targets. Of course, in everyday life we are more likely to see a headless human body than a headless grandfather clock. This result therefore potentially reflects some ambiguity with the detection of clock body-only targets. In other words, when searching for clock bodies, participants may have been unsure of exactly what they were searching for, leading to slower detection times for this target.

Taken together, the results of Experiment 1 are generally consistent with the notion that target size is related to target detection (i.e. bigger targets are processed fast than smaller targets). The results fail to provide any strong evidence in favour of a detection advantage for human faces, bodies and full-bodes over other comparable stimuli in natural (ecologically valid) scenes.

2.2 Experiment 2

2.2.1 Introduction

The findings Experiment 1 demonstrate no detection advantages for human targets in relation to the clock targets in natural scenes. In addition, the results appear to show that target detection is related to target size (Bindemann et al., 2010). Considering the extensive evidence suggesting that faces have an attentional advantage over other stimulus categories (e.g. Langton et al., 2008; Ro et al., 2007, Jenkins et al., 2002; Young et al., 1986), the lack of a human face advantage was unexpected.

Notwithstanding, it is possible that other factors are responsible for the lack of advantages for human targets in Experiment 1. For example, the content of target stimuli was blocked for target part and type in each block, i.e., only one target part (face, body, full-body) of one target type (human, clock) was presented per block. Participants could therefore always predict the target type and part on successive trials. Owing to this, it is possible that participants could have adopted different target detection strategies in each block, which could account for the differences in results between Experiment 1 and previous research (Bindemann et al 2010). Indeed, the type of experimental design and the way participants are presented with stimuli has been shown to influence how participants respond (e.g., Techentin & Voyer, 2007; Bruder & Ribeiro-do-Valle, 2009; Azevedo, Squella & Ribeiro-do-Valle, 2001). For example, Bruder & Ribeiro-do-Valle (2009) required participants to perform a go/no-go task, whereby participants' responded as fast as possible to a cross in a circle or vertical line in a circle (the go stimuli), and withhold a response when a circle within a circle (no-go stimulus) was presented. In the first experiment, they presented the targets in independent blocks of trials (go stimuli were always the cross or the vertical line in each independent block). The results demonstrated that there was no significant difference in reaction times to the cross or vertical line targets. However, in the second experiment they used a randomized design, whereby the target changed across trials in the block of trials (the target could either be the circle or the line). They observed that reaction times to the cross targets were significantly faster than reaction times to the vertical line targets. They argued that this difference reflected the cross targets being easier to discriminate than the line targets. They then suggested that to compensate for this difference in discriminability, when the targets were presented in independent blocks where participants could predict each successive trial, participants adopted an attentional set to process the vertical line targets.

An attentional set is a specific type of search strategy that prioritizes stimuli for selection on a number of simple features. Top-down processing favours task-relevant information and the demands of the task determine the adopted strategy. Therefore, it is thought that observers will always choose the attentional set that they believe will maximize their performance (Leber & Egeth, 2006). For this reason, Bruder and Ribeiro-do-Valle (2009) suggested that to optimize search for the line targets, when presented in independent blocks, participants adopted such an attentional set whereby they deployed more intense attention to the line targets and less intense attention to the cross targets. The adoption of such a set consequently masked reaction time advantages for the cross targets (Bruder and Ribeiro-do-Valle, 2009). They argue that adopting a similar set in the randomised blocks could not occur because participants cannot predict whether a cross or vertical line target would be presented next.

Therefore, presenting the stimuli in independent blocks, as was the case in Experiment 1, could account for the lack of detection advantages for the human targets. Specifically, it is possible that clock targets are less discriminable (possibly because they are more ambiguous) than human targets. Thus, given that in Experiment 1 the targets were always presented in independent blocks, to compensate for the difficulty in discrimination, participants may have adopted a specific compensatory attentional set to detect the clock targets. Namely, more attention was deployed to clock parts, subsequently masking any potential detection advantages for the human targets.

To address this design issue, Experiment 2 used a randomized design. In which human and clock targets were paired up on the basis of target part only. Participants were now presented with three blocks of trials: face-only, body-only, and full-body. This enhanced the unpredictability of the target type, reducing the by possibility of participants adopting a specific attentional set for detecting clock targets. It was predicted that if blocking trials in Experiment 1 facilitated participants in adopting an attentional set, which masked advantages for the human targets, then randomizing human and clock targets on a trial-by-trial basis should reveal advantages for human targets. This finding would suggest that the lack of advantages for human targets in Experiment 1 could be attributable to the use of a blocked design.

2.2.2 Method

2.2.2.1 Participants:

A convenience sample of 39 participants (31F, 8M) between 18-23 years was taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 30 minutes.

2.2.2.2 Design, Stimuli, Apparatus, Materials and Procedure

The apparatus and background stimuli were exactly the same as in Experiment 1. The experiment was run in three blocks: face-only, body-only, and full-body. In each block, target type (human, clock) was randomized from trial-to-trial. All blocks were preceded by 8 practice trials and 120 experimental trials comprising of 72 target-present trials (36 target human trials, 36 target clock trials) and 48 target-absent trials. The timing of the presentation of the stimuli was exactly the same as Experiment 1. The ordering of blocks and trials within blocks was randomized for every participant. In all other respects, the design and procedure were identical to those of Experiment 1.

2.2.3 Results

As in Experiment 1 incorrect button presses and none responses were counted as errors, these were removed before analysis. The average of the participants' median RTs in each of the conditions are shown in Figure 9, along with the mean percentage of errors. As can be seen from Figure 9, in general, participants made more errors detecting clock targets (3.3% errors) than human targets (1.6% errors). As is clear from Figure 9, the error rates mirror that of the RT data, indicating that there is no trade-off between speed and accuracy, as such, error rates are not further examined here.

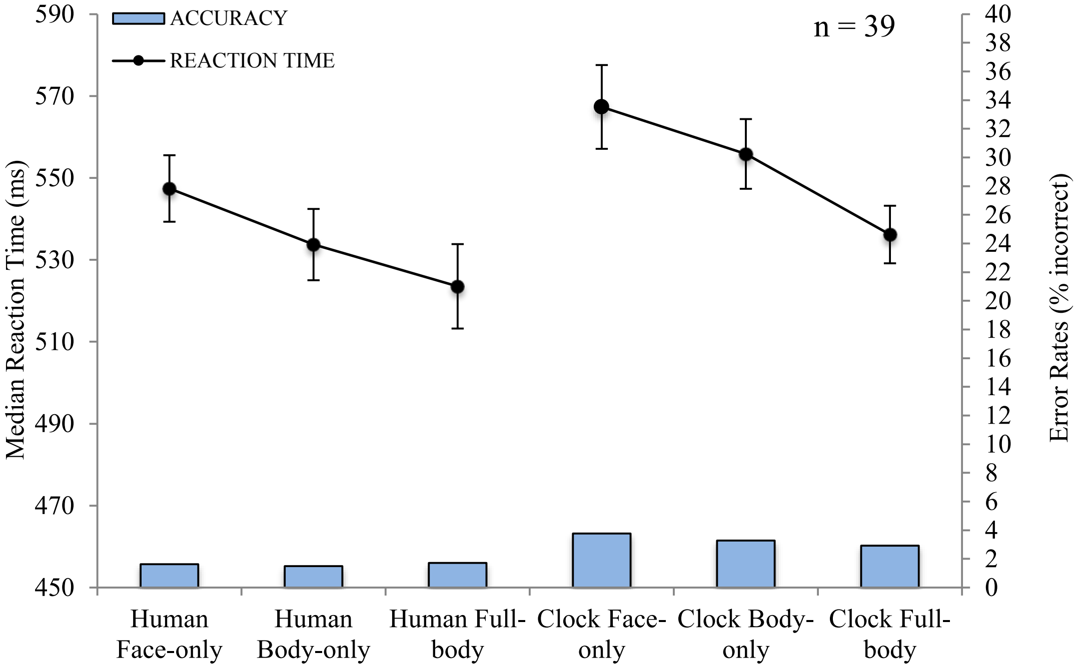


Figure 9. Means of the median correct detection rates and error rates for each condition of Experiment 2 for the target present conditions. The vertical error bars represent 95% within subject confidence intervals (Cousineau, 2005; Morey, 2008).

A two-way within subjects ANOVA with factors of target type (human, clock) and target part (face-only, body-only, full-body) was conducted on the RT data. This analysis revealed a significant main effect of target type, $F(1, 38) = 26.72$, $MSE = 19434.00$, $p < 0.01$, $\eta^2 = 0.06$, indicating that human targets (M=534.90) were detected faster than clock targets (M=553.13). There was also a significant main effect of target part, $F(2, 76) = 5.04$, $MSE = 14783.05$, $p < 0.01$, $\eta^2 = 0.09$, reflecting significantly faster target detection for the full-body targets (M= 529.87) compared to the face-only (M=557.37, $d = 0.33$), and body-only targets (M=544.82, $d = 0.18$). Target detection for body-only targets (M=544.82) was also significantly faster than detection of face-only targets (*M*=557.37; all $p < 0.05$, $d = 0.15$). The interaction between factors target type and part interaction was not significant, $F(2, 76) = 1.02$, $MSE = 479.693$, $p > 0.05$, $\eta^2 = 0.00$.

2.2.4 Discussion

Inconsistent with Experiment 1, the human targets were detected faster than the grandfather clocks over all. Instead, this finding is consistent with the idea that human body/body targets are processed faster than other comparable objects/object parts when presented in natural scenes. In line with the hypothesis, one reason that this was not observed in Experiment 1 was that the human target advantage was masked by the way the target objects were presented. More specifically, the way in which the target objects were presented in independent blocks led to participants adopting an attentional set that equalled out target type detection. However, it is also possible that some other factor (e.g., low-level image-based properties) may have resulted in an apparent but not real attentional advantage for the human targets. This effect would therefore benefit from further investigation and is explored in Experiment 3.

The results also demonstrate that irrespective of type (human, clock), full-body targets had a detection advantage over both face-only and body-only targets, and body-only targets had an advantage over face-only targets. The significant effect of part is consistent with Experiment 1 and supports the notion that as target size increases time to target detection decreases, irrespective of target type (Proulx & Egeth, 2008; Proulx, 2010).

2.3 Experiment 3

2.3.1 Introduction

Experiment 2 provided support for the notion that human faces, bodies and full-bodies have attentional advantages in natural scenes, showing detection advantages for human targets relative to clock targets. Whilst this effect was observed, it was inconsistent with the results of Experiment 1. Although this difference could be attributable to a variation in stimuli presentation, it is also possible that low-level image-based properties made the human targets more salient in the scenes, facilitating faster detection (Langton et al., 2008). Experiment 3 aimed to explore the apparent advantages for the human targets observed in Experiment 2. For this purpose, Experiment 2 was replicated, using inverted (rotated 180 degrees) targets.

As discussed in Chapter 1 (Section 1.5), one of the most robust and well-established markers for the special processing of faces and bodies is the behavioural face and body inversion effect. Namely, inverting a face or body leads a relative drop in detection accuracy and target response times as compared to upright faces and bodies (Yin, 1969; Damasio et al., 1982; Reed et al., 2003). Importantly, what the inversion effect demonstrates is that inverted faces and bodies do not attract the allocation of attention. As such, the presence of a face or body inversion effect can be used as a diagnostic for an attentional priority towards upright faces.

An important condition of the inversion effect is that when faces and bodies are inverted, the low-level image-based properties (e.g., local contrast, edge density, luminance, spatial frequency) that might make a face or body salient in a scene remain constant, yet detection and recognition is impaired. Accordingly, in Experiment 3 detection performance was compared between inverted clocks target parts and inverted human target parts. It was predicted that if the detection advantages for human targets relative to clock targets observed in Experiment 2 were a result of low-level confounds then advantages for human targets should be preserved when inverted.

In addition, detection performance was also compared between the upright human and clock target parts (from Experiment 2) and inverted human and clock target parts

(Experiment 3). Given that the behavioural body inversion effect strongly indicates that inverted bodies do not attract the allocation of attention, it was reasoned that if observed detection advantages for the upright human targets in Experiment 2 arose because of some attentional processing advantage, detection rates should be significantly faster for upright human targets versus inverted human targets. However, if some other factor was at play in Experiment 2 then there should be no difference between any of these target comparisons. Given that grandfather clocks should not be subject to special processing in the visual system, it was predicted that there would be no difference between the detection rates of upright clock targets (Experiment 2) and inverted clock targets (Experiment 3).

2.3.2 Method

2.3.2.1 Participants:

A convenience sample of 46 participants (40F, 6M) between 18-27 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 30 minutes.

2.3.2.2 Apparatus, stimuli, procedures, and design:

The design, apparatus, and procedures were all identical to Experiment 2. The background scene stimuli and all targets (including practice trials) were all rotated through 180 degrees to produce the inverted stimuli in Experiment 3. The results from Experiment 2 for the upright human and clock targets were used for the upright/inverted analysis in the current experiment.

2.3.3 Results

Again incorrect button presses and none responses were counted as errors, these were removed before analysis. The means of the participants' median RTs in each of the conditions for Experiment 3 are shown in Figure 10, along with the mean percentage of errors. As can be seen from Figure 10, once again error rates were low, participants made errors for just 2.12% of the human targets and 3.42% of the clock targets. As can be seen from Figure 10, the error rates behave in a similar way to the RT scores, with more errors in conditions that participants performed the slowest in. However, there appears to be no evidence of a speed accuracy trade-off.

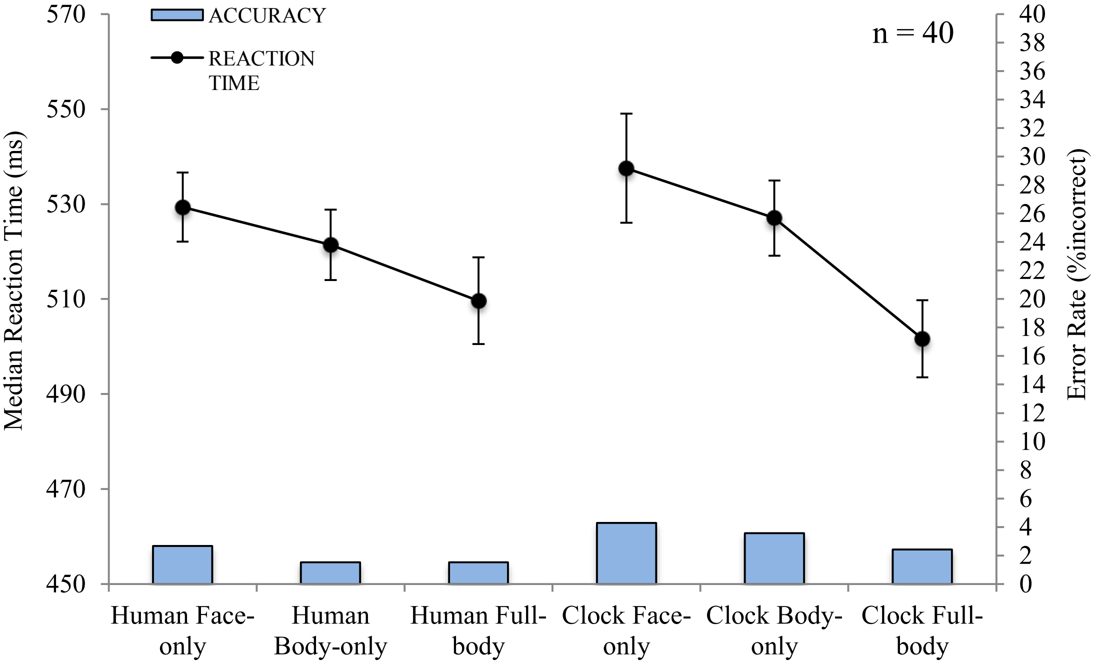


Figure 10. Means of the median correct detection rates and error rates for each condition of Experiment 3 for the target present conditions. The vertical error bars represent 95% within subject confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 × 3 × 2 within subjects ANOVA with factors of target type (human, clock), target part (face-only, body-only, full-body), and orientation (upright, inverted) was conducted on the RT data. This analysis yielded a significant main effect of target type, $F(1, 83) = 18.47$, $MSE = 12859.34$, $p < 0.01$, $\eta^2 = 0.02$, and a significant main effect of target part, $F(2, 166) = 10.45$, $MSE = 32778.49$, $p < 0.01$, $\eta^2 = 0.09$. The was however, no main effect of orientation, $F(1, 83) = 1.51$, $MSE = 6646.16$, $p > 0.05$, $\eta^2 = 0.00$. The interaction between target part × target type was significant, $F(2, 166) = 3.412$, $MSE = 1913.02$, $p < 0.05$, $\eta^2 = 0.01$. There was also a significant interaction between target type × orientation, $F(1, 83) = 12.08$, $MSE = 8410.36$, $p < 0.01$, $\eta^2 = 0.01$. There was no significant part × orientation interaction, $F(2, 166) = 0.55$, $MSE = 173.17$, $p > 0.05$, $\eta^2 = 0.00$. There was no significant type × part × orientation interaction, $F(2, 166) = 0.38$, $MSE = 210.37$, $p > 0.05$, $\eta^2 = 0.00$.

To examine the interaction between target type and target part, corrected pairwise comparisons were used. These revealed that regardless of orientation, for human target parts, target detection was significantly faster for the full-body targets ($M = 516.03$) than for face-only ($M = 537.65$; $p < 0.05$, $d = 0.36$) and body-only ($M = 527.06$; $d = 0.19$) targets. There was a significant difference in the detection rates of body-only targets ($M = 527.06$) and face-only targets ($M = 537.65$), reflecting the faster detection of the body-only targets ($p < 0.05$, $d = 0.20$).

For the clock targets, regardless of orientation, full-body targets ($M = 517.48$) were detected significantly faster than the body-only ($M = 540.26$; $p < 0.05$, $d = 0.45$) and face-only ($M = 551.21$; $p < 0.05$, $d = 0.49$) targets. Although body-only targets ($M = 540.26$) were detected numerically faster than the face-only targets ($M = 551.21$), there difference did not reach significance ($p > 0.05$, $d = 0.17$). All other significant comparisons were not relevant to the hypotheses and are therefore reported in Appendix 1.

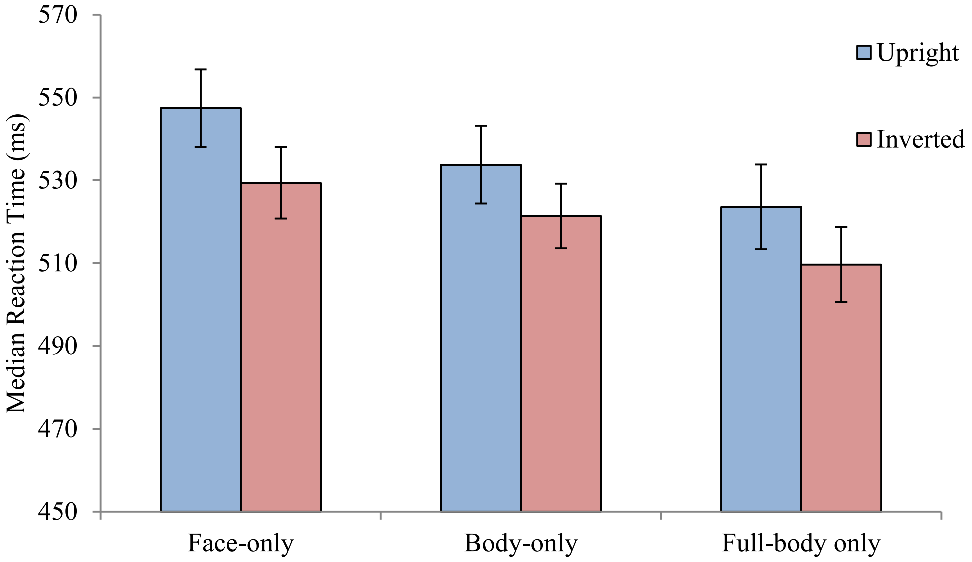


Figure 11. Means of the median correct detection rates for upright human target parts (Experiment 2) compared to inverted human target parts (Experiment 3). Vertical bars presented the between subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008).

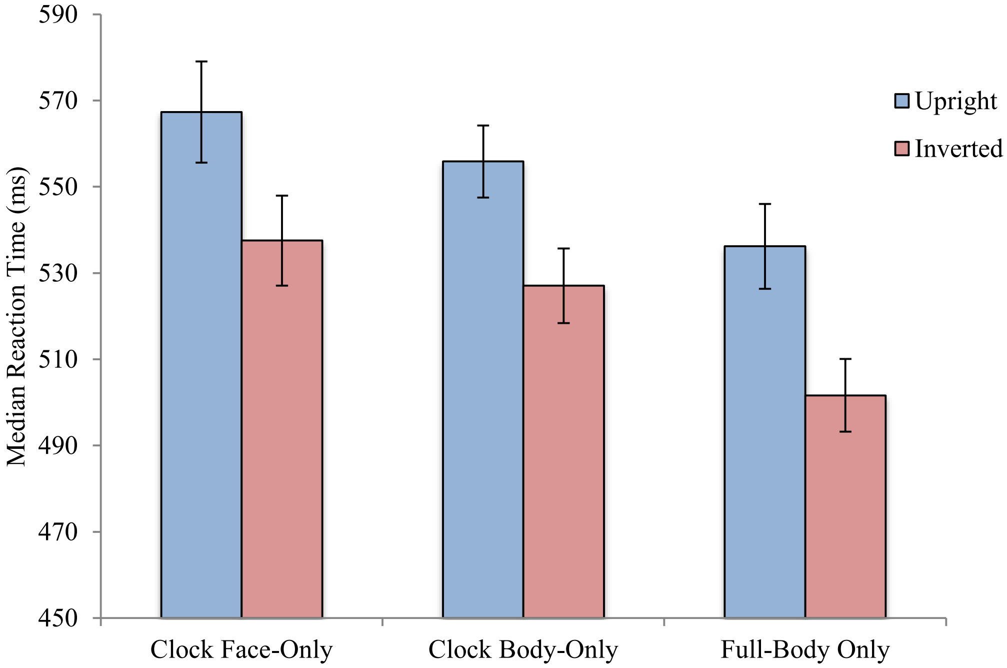


Figure 12. Means of the median correct detection rates for upright clock target parts (Experiment 2) compared to inverted clock target parts (Experiment 3). Vertical bars presented the between subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008).

The means of the participants' median RTs for the upright human target parts (Experiment 2) compared to the inverted human target parts (Experiment 3) is shown in Figure 11. Additionally, the means of the participants' median RTs for the upright clock target parts (Experiment 2) compared to the inverted clock target parts (Experiment 3) are shown in Figure 12. To examine the interaction between target type and orientation, corrected independent t-tests revealed that there was no significant difference between the detection rates of the inverted human targets from Experiment 3 ($M = 520.14$) compared to the upright human targets from Experiment 2 ($M = 534.90$; $p > 0.05$, $d = 0.17$). There was also no significant difference between the detection rates of the inverted clock targets from Experiment 3 ($M = 522.07$) and upright clock targets from Experiment 2 ($M = 553.13$; $p > 0.05$, $d = 0.36$). Paired comparisons revealed that there was no significant difference between the detection rates of the inverted human targets from Experiment 3 ($M = 520.14$) and the inverted clock targets from Experiment 3 ($M = 522.07$; $p > 0.05$, $d = 0.09$). They also revealed that upright human targets from Experiment 2 ($M = 534.90$) were detection significantly faster than upright clock targets from Experiment 2 ($M = 553.13$, $d = 0.23$).

2.3.4 Discussion

Experiment 3 explored the effect of inverting the target stimuli on detection times to the targets. The results demonstrated that inverting the human and clock target parts eliminated the detection advantage for human targets over the clock targets observed in Experiment 2. In line with the hypothesis, since the stimuli were identical apart from orientation, the detection advantage for the human targets observed in Experiment 2 were most likely not due to low-level image-based properties. Instead, this result might suggest that the human target detection advantages in Experiment 2 are attentional based (e.g., Langton et al., 2008; Ro et al., 2007; Downing et al., 2004).

On the other hand, the comparison of the detection rates between the upright human target parts from Experiment 2 and the inverted human target parts from Experiment 3 revealed no detection advantages for upright human targets compared to inverted human targets. As discussed previously, the presence of an inversion effect for faces and bodies has been taken as evidence that inverted faces and bodies do not attract attention and that the processing of upright faces and bodies is special, such that they capture a viewer's attention (Reed et al., 2003, 2005; Ro et al., 2007). Therefore, given that the inversion effect was not observed for any of the human target parts, the current findings strongly suggest that attention is not being attracted or captured by the upright human targets. Supporting this, there was also no significant difference observed between the detection rates of the upright and inverted clock target parts. Observing that both the human and clock targets do not demonstrate inversion effects suggests that they may be processed in the same way. Therefore, these findings strongly indicate that the detection advantages for human targets in Experiment 2 cannot be attributable to some inherent attentional processing advantage for human faces, bodies and full-bodies. Rather these findings support the view that some other factor produced these advantages.

Another interesting finding worthy of discussion from Experiment 3 was that consistent with Experiments 1 and 2, for both human and clock targets, the full-body targets were detected faster than the body-only and face-only targets, and the body-only targets were detected faster than the face-only targets, regardless of orientation. These findings support the notion that faster reaction times are associated with increased target size and not because they are special objects, though this requires further investigation.

2.4 General Discussion

The three experiments reported in this Chapter have produced a number of interesting findings. Previous research (Bindemann et al., 2010) reported similar detection rates for human faces and bodies in natural scenes, suggesting that they might be processed in a similar way. Experiment 1 aimed to replicate this finding and extend it by comparing detection times for these targets to analogous clock targets. However, the results from Experiment 1 demonstrated that bodies were detected faster than faces, possibly reflecting a relationship between target detectability and target size, rather than a comparable processing advantage for these targets. Importantly, the results also revealed that there were no detection advantages for human targets relative to the grandfather clock targets. However, Experiment 1 presented stimuli in a blocked design, which potentially masked advantages for human targets. Experiment 2 used a randomized design to address this issue. Experiment 2 demonstrated detection advantages for the human targets relative to the clock targets. Therefore the failure to observe advantages for human targets in Experiment 1 could potentially be explained in terms of the use of a blocked design. These advantages were eliminated when all targets in the scenes were inverted in Experiment 3; ruling out an explanation for the advantages seen for the human targets based on low-level image properties. However, when the detection rates for the inverted human targets from Experiment 3 were compared to detection rate for upright targets from Experiment 2, no advantage for the upright targets was observed. This suggests that some factor, other than an inherent attentional processing advantage for human targets, facilitated the detection advantage in Experiment 2.

Further, inconsistent with Bindemann et al. (2010), across Experiments 1 – 3 and regardless of the orientation of the human and clock targets, the results were consistent with the larger full-body targets being detected faster than both the smaller face-only and body-only targets, and body-only targets being detected faster than face-only targets. It therefore appears that the relative salience of the physically larger targets played an important role in driving the detection advantages of the larger targets relative to the search for the smaller targets (Lamy, Amunts, & Bar-Haim, 2008). This is in line with previous research that large objects in a scene capture attention, independent of top-down goals (Proulx & Egeth, 2008; Proulx, 2010).

The absence of an inversion for the face-only targets is particularly striking, considering the volume of research in this field strongly suggesting that the inversion effect is a defining characteristic for face processing (Yin, 1969; Reed et al., 2004; Diamond & Carey, 1986; Farah, Tanaka, & Drain, 1995; McKone et al., 2007; Yin, 1969). This finding was unexpected, however, the face inversion effect may depend upon the task and the nature of the stimuli used. Specifically, the evidence for the face inversion effect in face detection tasks that use natural backgrounds has been mixed. As such, the lack of an inversion effect here may not be sufficient enough criteria by which to rule out attentional processing advantages for these stimuli in Experiment 2. (Lewis & Edmonds, 2003, 2005; Rousselet, Mace & Fabre-Thorpe, 2003; VanRullen, 2006; Garrido et al., 2008; see General discussion, section, 6.2. for a full discussion on this point).

In summary, the experiments in Chapter 2 have shown a number of things. First, they have established that Bindemann et al.'s (2010) finding that bodies are detected as fast as faces is not replicable. Instead, it appears target detectability is related to target size, with faster detection times for larger targets (Bindemann et al., 2010). Second, that there is no strong evidence in favour of the hypothesis that human faces, bodies or full-bodies have some kind of attentional processing advantage that allows them to be processed more efficiently than comparable objects, at least in complex natural scenes. However, given the inconsistencies in the current results, the question whether human faces, bodies and full-bodies have attentional advantages in natural scenes would benefit from further investigation and is explored in Chapter 3, using an alternative paradigm.

Chapter 3

3 Attention Capture by Human Faces and Bodies Using the Additional Singleton Paradigm

3.1 Introduction

The experiments in Chapter 2 used a visual search paradigm to replicate previous findings by Bindemann et al. (2010) and examine the hypothesis of whether human faces, bodies or full-bodies have an attentional processing advantage over other objects, when presented in natural backgrounds. Experiment 1 showed that there were no apparent detection advantages for the human targets, whereas Experiment 2 revealed detection advantages for human targets, suggesting that experimental design may have masked advantages in Experiment 1. Experiment 3 revealed that advantages for human targets could not be explained in terms of low-level image-based confounds. However, a comparison of the detection rates between the upright human and clock targets from Experiment 2 and inverted human and clock targets from Experiment 3 revealed no inversion effect for any of these targets, suggesting that detection advantages for the human targets are not because of some special attentional processing advantage.

Although attention researchers have frequently and fruitfully used the visual search paradigm, it is possible that it is not the most suitable paradigm for determining attention capture per se. In this Chapter, we continued to explore apparent detection advantages for human faces, bodies and full-bodies, but using an alternative methodology to the visual search paradigm. Namely, the additional singleton paradigm, which is a more stringent test of attention capture (Theeuwes, 1991, 1992, 1994, 1996). In the standard visual search task, targets that are detected immediately and without effort are assumed to pop out of the scene. Such an effect is used as a diagnostic for determining whether that target is detected preattentively, in the parallel stage of visual search. This process has been taken to suggest that the target is receiving attentional prioritization independent from the intentions of the observer (Theeuwes, Kramer, Hahn, Irwin & Zelinsky, 1999). In other words, attention is captured in a bottom-up manner by the target. However, just because a target can be detected fast and efficiently, does not necessarily indicate attention capture by this target (Yantis & Egeth, 1999).

Importantly, in a visual search task, participants are given the task of searching for a pre-specified target. However, it is well-established that top-down processing favours task-relevant information, and the demands of a task will determine the strategy (attentional set) that a viewer will adopted to search for a target in a scene (Arnott, Pratt, Shore, & Alain, 2001; Folk, Remington, & Johnston, 1992; Folk, Remington, & Wright, 1994; Gibson & Kelsey, 1998; Lambert, Wells, & Kean, 2003; Remington, Folk, & McLean, 2001). For example, when a viewer is searching for a red flower amongst a bouquet of various coloured flowers, the red flower will be detected more easily than the other flowers. In this case, participants have readied themselves with the intention to search for that target and adopted a top-down strategy to attend the target one is looking for (Folk et al., 1992; Folk, Remington & Johnston, 1993). However, as noted by Yantis and Egeth (1999) attention capture by a target can only occur when the target is completely task irrelevant and observers have no incentive to attend it intentionally. Only when a target captures attention under these conditions can it be said to capture attention in a stimulus-driven/bottom-up fashion (Yantis & Egeth, 1999). Therefore, any performance benefit for a target in the visual search paradigm cannot be used to indicate whether a target has captured attention in a bottom-up manner (Becker, 2007). Rather detection advantages found when using the visual search paradigm most likely represent processing advantages, rather than attention capture.

In order to directly assess bottom-up attention capture, researchers have used modified versions of the visual search paradigm, using the notion of task-irrelevant singletons. One of the most commonly used methods of this kind is the additional singleton paradigm (ASP), introduced by Theeuwes (1991, 1992, 1994, 1996). In the classic ASP, observers are asked to perform a visual search for a unique target amongst an array of homogeneous distractors. According to theories of visual search this type of search should be done in parallel (Triesman & Gelade, 1980), or efficiently (Wolfe, 1994). However, in the ASP, on some trials an additional salient distractor, which is task irrelevant, is also present somewhere else in the scene. Reaction times for target detection in scenes that contained both a target and distractor are then compared to a condition in which the distractor is not present. For instance, in one of Theeuwes (1992) experiments, participants searched for a green circle surrounded by green squares, participants then determined the orientation of a line presented in the target circle. In the distractor condition, a red square distractor was also present in the display (see Figure 13. for example). It was observed that target search increased as a function of the distractor being present. Theeuwes (1991, 1992, 1994, 1996) argued that this slowing in reaction time to the target in the presence of the distractor could be explained in terms of attention capture. This is because attention was automatically captured by and orientated towards the irrelevant singleton, therefore taking longer for attention to be reoriented towards the actual target. Specifically, during serial search of a target-present array consisting of between 1 and *n* items, on average (n +1)/2 items need to be searched before the target is found. Therefore, if a task irrelevant distractor captures attention first, on average more of the scene would then need to be searched in order to locate the target. Conversely, when the task irrelevant distractor is absent, on average only half of the scene will be searched before the target is located, resulting in shorter target detection times (Langton et al., 2008).

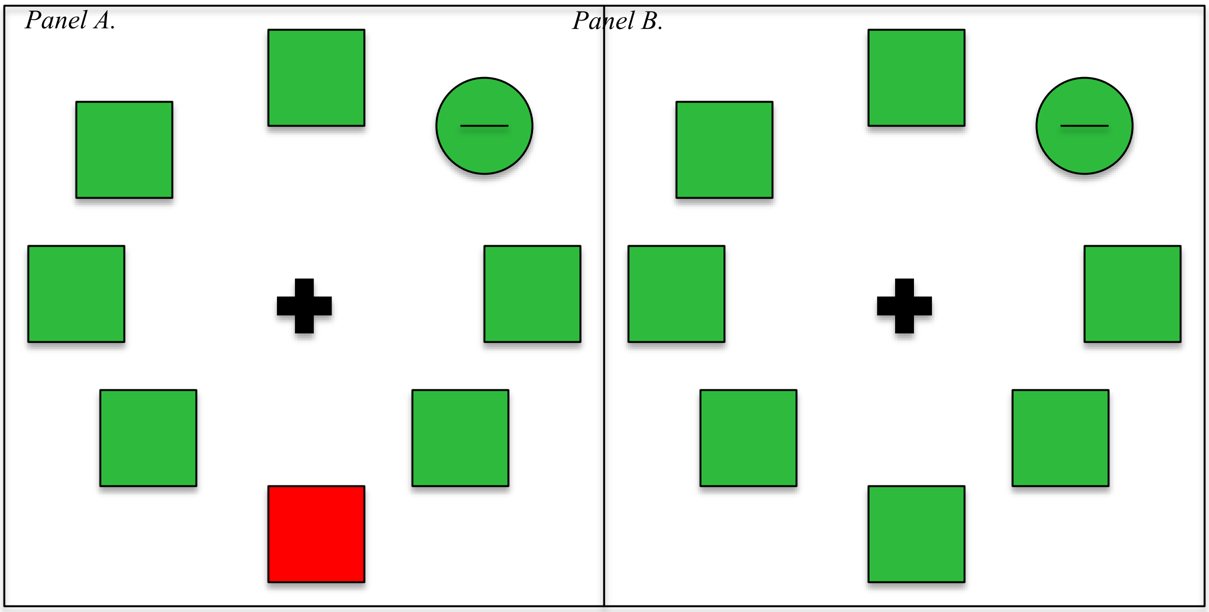
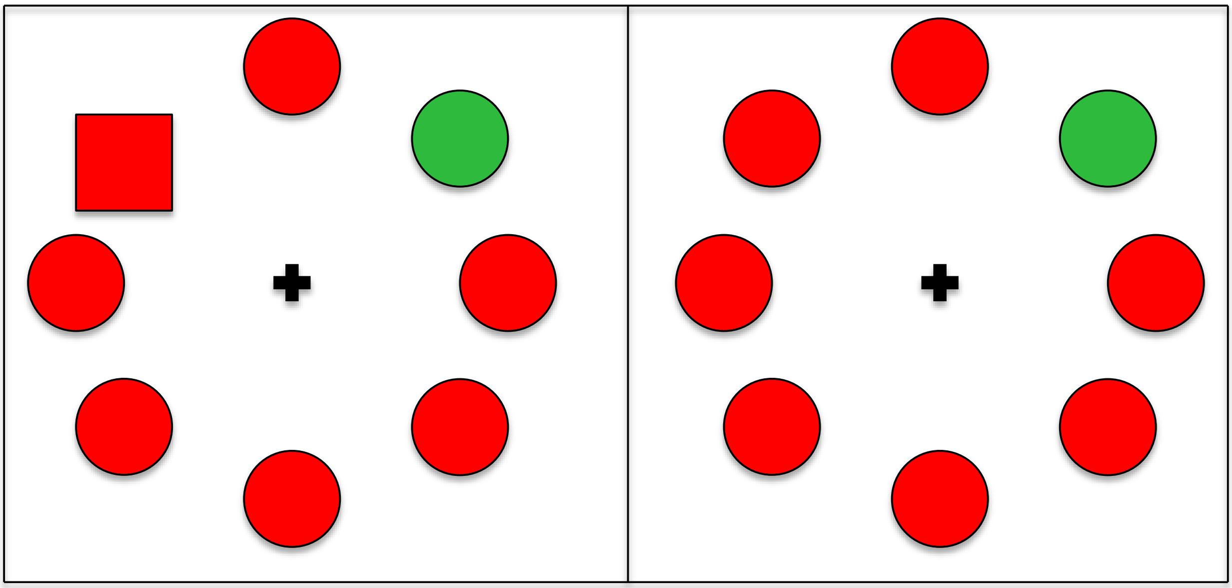


Figure 13. A recreation of Theeuwes (1992) arrays. Panel A shows an example of an array containing both a target (green circle) and a distractor (green square). Panel B shows an example of an array containing only a target (green circle).

Important to note is that in Theeuwes (1992) subsequent experiments when the target was a green circle among red circles, a red square distractor did not influence the search for the target (see Figure 14. for example). From this, Theeuwes concluded that the selection process depends on bottom-up processing, whereby the selection of the target is contingent on the relative salience of the surrounding items. Specifically, using this example, a green circle target amongst red circle targets produces a higher activation than the distractor red square in an array of red circles. Hence attention is generally guided to the location that shows the largest activation, if this is not the target, attention is involuntarily moved on to the next area of most salience (c.f. Cave & Wolfe, 1990). In the current example the most salient item is the different coloured distractor and then the different shaped target. This explains why only the colour distractor, not the different shaped distractor, slows responses to the target. Further, Theeuwes (1992) suggested that preattentive selection processes are determined by bottom-up influences rather than top-down processes. Importantly, what this suggests is that when using the ASP, regardless of top-down goals of the observer (i.e. searching for a specified target), the most salient object in the scene will preattentively capture attention in an involuntary, stimulus-driven manner. Another major advantage of this paradigm is that the distractor is *never* the target and can therefore play no role in the observer's search strategy (attentional set). Consequently, this paradigm fulfills Yantis and Egeth's (1999) criterion for stimulus driven attention capture, as the distractor is both task-irrelevant and outside of the viewer's intentions (not the target of search).



Panel A. *Panel B.*

Figure 14. Panel A is an example of an array with a target (green circle) and distractor (red square). Panel B is an example of an array containing a target (green circle). Theeuwes (1992) argued that the target green circle amongst green squares produces a much lower activation than the target green circle among red circles.

Nevertheless, only a few studies have applied the ASP to examine whether human faces and bodies can preattentively summon attention. Langton et al. (2008) examined whether task-irrelevant faces slow reaction times to a target. Using a version of the ASP, participants were required to search an array of six items with the task of detecting a target butterfly. On half of the trials a distractor face was also present, on the other half of the trials it was absent. They found that the inclusion of the face significantly slowed responses to the target compared to when it was absent. This face distractor effect was consistent with paradigms notion that selective attention had been automatically directed to the face, ensuring that it was examined first in the display. As Langton et al. (2008) acknowledged, it was possible that low-level image properties, such as edge density and local contrast, might have made the distractor face more salient within a display thus facilitating attention capture. In order to control for this possibility, they inverted all their stimuli. However, the distractor face effect was removed when the items were inverted, suggesting that a higher-level representation of faces, rather than low-level image properties had produced the distractor face effect. Langton et al. further argued that because both the butterflies and face were the only objects in the array that shared the property of animacy, this could have resulted in attention capture. In a following experiment, they presented faces as the targets and butterflies as the distractors. They proposed that if faces were capturing attention because of the shared property of animacy, butterflies should also capture attention when a face is the target. In addition, they also argued that given that faces were present on only half the trials, it might have been that the novelty of their appearance caused them to capture attention. Therefore they argued that if novelty had caused attention capture by faces, when butterflies were novel they too should capture attention. Despite these manipulations, the distractor butterfly targets did not slow responses to the face targets, suggesting that neither the shared property of animacy nor novelty had caused the faces to capture attention but rather that the distractor face effect was mediated by an attentional bias (for the faces). Indeed, Langton et al. suggest that these data are in line with models of visual search, such as Guided Search (Wolfe, 1994), whereby pre-attentive processes guide attention to items in order of salience.

Comparable attentional effects have also been observed with human body parts. As discussed in Chapter 1 (see section 1.4.3. for full review), Ro et al. (2007) used a variation of the ASP to examine search performance and distractor effects for human faces, body parts and other object categories. They presented all stimuli in outline colour frames and defined both targets and distractors on this basis. Targets were presented in green frames; distractors in red frames and all other items were presented in blue frames. Participants searched for the green frame amongst the blue frames and made a categorical judgment on the item in the frame based on a preceding word cue. On half of these trials a distractor red frame was also present. The use of the colour frames allowed them to exploit the concept of a task-irrelevant colour singleton capturing attention (see, Theeuwes, 1994, 1996). Specifically, as is known from visual search (Treisman & Gelade, 1980), a colour singleton will capture attention. Thus, in an array of blue frames, the target green frame and distractor red frame will automatically orientate the viewer's attention to the object within the singleton frame. This manipulation therefore allowed Ro et al. to examine how faces and body parts were processed once attention was allocated to the location of the stimulus of interest. They observed that when faces and body parts were presented in the target green frames, search was faster compared to when non-face and non-body objects were within the green frame. Importantly, they observed that when a face or body part was in the distractor red frame, response times to the target green frame were significantly slower, compared to when a non –face or –body part was present in the red frame, indicating that there may be an attentional bias towards human faces and bodies. However, as Ro et al. acknowledge, given that the singleton colour frames captured attention, this study does not demonstrate that they capture attention per se, rather it demonstrates that when attention is allocated to the body or face, they may be automatically processed.

In summary, evidence from Langton et al. (2008) and Ro et al. (2007) appears to suggest that task-irrelevant human faces and bodies may capture attention and be processed more efficiently, respectively. However, in both these studies, objects were presented in artificial grey backgrounds and as part of an array. As discussed in previous chapters, observed advantages for faces and body parts are therefore confined to these displays and cannot be generalized to more real-world circumstances. The aim of the next set of experiments was to provide a further test for attention capture by human faces and bodies in both natural and grey scenes. To address these concerns, the current experiments used a variation of the ASP (Theeuwes, 1991, 1992, 1994, 1996). Of specific interest was whether using this paradigm would reveal any evidence for faces and bodies capturing attention in a stimulus-driven bottom-up manner, when they are task-irrelevant. If human faces and bodies capture attention stimulus-driven bottom-up manner, then search times for displays that contained a face or body should be slower compared with displays where no human face or body was present.

3.1.1 Experiment 4

3.1.1.1 Introduction

Experiment 4 explored whether human faces are able to capture attention under conditions where they are completely unrelated to the goal of the experiment, i.e., task-irrelevant (Yantis & Egeth, 1999). In order to test this, a variation of the ASP was used. In contrast to previous studies using the ASP (e.g., Langton et al., 2008; Ro et al., 2007), here stimuli were presented in natural scenes rather than arrays of objects. A condition of the ASP is that attention capture can be inferred if it can be demonstrated that a task-irrelevant singleton has had a detrimental effect on search time for the target. Therefore, in Experiment 4, participants were required to search natural scenes for a non-face target (a ladybird). On half of these trials a task-irrelevant distractor (a human face) was also present in the scene. If faces capture attention, then target detection will be significantly slower when target search takes place in the presence of an accompanying face relative to when a face is absent (referred to as the face distractor effect).

Furthermore, it was also possible that advantages for human faces arise because of some experimental artefact owing to use of natural backgrounds for the presentation of stimuli. To explore this issue, it was also necessary to present the stimuli under conditions that eliminate this possibility, i.e. by presenting the stimuli in uniform grey backgrounds. Therefore all participants completed the experiment in both natural and uniform grey backgrounds. It was predicted that if advantages for human targets only occur because of some artefact of the natural backgrounds, then when presented in the grey backgrounds a task-irrelevant face should not slow the search for the target. The use of the grey backgrounds also had a number of other experimental advantages. First, given that most previous research in face detection has used uniform colour backgrounds to present their stimuli, using grey scenes also replicates previous research methodologies. Therefore, the observation of a face distractor effect in the grey scenes would provide support for attention capture by faces being a robust effect. Second, as both target and distractor were presented on uniform grey backgrounds, they differed from the background in a number of features, e.g. colour, luminance, form etc., making them highly salient with respect to the grey backgrounds. Presenting both target and distractor in the grey scenes was therefore also a measure of how much stimuli captured attention because of low-level factors. It was predicted that if low-level factors were facilitating attention capture, then search in the grey scenes should be faster than search in the natural scenes.

3.1.1.2 Method

3.1.1.2.1 Participants

A convenience sample of 30 participants (25F, 5M) between 18 – 23 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 20 minutes.

3.1.1.2.2 Design

A 2 × 2 × 2 within–subjects design was used with three factors: target (present vs. absent), distractor (present vs. absent) and background (natural vs. grey). The dependent variables were the reaction time taken from stimulus onset and participant response and participants accuracy. The experiment used an ASP paradigm and the data were analysed using a 2 × 2 × 2 analysis of variance. Post hoc analysis (where appropriate) used paired *t* tests with a false-discovery rate correction for multiple comparisons (Benjamini and Hochberg, 1995).

3.1.1.2.3 Stimuli

The 24 natural backgrounds (12 inside, 12 outside) and 12 face stimuli used in Experiment 4 were identical to those used in Chapter 2. The backgrounds used were reduced to 24 from the original 30 backgrounds to fit with the experimental design. The target ladybirds were created from 12 different pictures of ladybirds. Ladybirds were chosen as the target as they are highly salient and familiar objects. To create the targets, they were all cropped out of their natural outline and manipulated to be approximately the same size as one another. The grey backgrounds and all other manipulations were completed using ADOBE Photoshop (CS 8). For the target present scenes, either the target was presented in isolation with no distractor present (target-present/distractor-absent condition), or the target and distractor were presented (target-present/distractor-present condition). For the target absent scenes, either the distractor was presented in isolation (target-absent/distractor-present condition), or neither the distractor nor the target was present (target-absent/distractor-absent condition). Example stimuli illustrating the experimental conditions can be seen in Figure 15.

Natural target-absent/ distractor-absent

Natural target-absent/ distractor-present

Natural target-present/ distractor-absent

Natural target-present/ distractor-present

Grey target-absent/ distractor-absent

Grey target-absent/ distractor-present

Grey target-present/ distractor-absent

Grey target-present/ distractor-present

Figure 15. An illustration of the stimuli for each of the experimental conditions in each phase of Experiment 4.

3.1.1.2.4 Apparatus and Materials

The apparatus and materials were identical to those used in Experiments 1 – 3 (Chapter 2).

3.1.1.2.5 Procedure

The experiment took part in two phases. In phase one, participants searched for the targets in natural backgrounds. In phase two, participants searched for the targets in grey backgrounds. The procedure was identical in both phases; however, the phase completed first was counterbalanced across all participants. Every trial began with the onset of a fixation cross which was displayed for 1000 milliseconds (ms). On each trial the location of the fixation cross was randomly rotated around four possible onscreen locations, corresponding to the four corners of the screen. Immediately after this, the scene stimulus was displayed until a target present/absent response had been made. If no response was made within 3,000ms, the scene was removed and the next trial was initiated. Participants were asked to quickly and accurately determine whether a target was present or absent in a scene and to make their responses on the button box, using two fingers from their preferred hand. Each phase, natural/grey, comprised of 144 experimental trials, consisting of 72 target present scenes (36 trials for target-present/distractor-absent scenes, 36 trials for target-present/distractor-present scenes) and 72 target-absent scenes (36 trials for target-absent/distractor-present scenes, 36 trials for target-absent/distractor-absent scenes).

In the natural phase of Experiment 4, on every trial the experimental software (Psychopy) selected a background at random. Each background was used an equal number of times. In both phases of the experiment, in target present scenes, Psychopy randomly selected a target. Similarly, in distractor present scenes, distractors were also selected at random. Over the course of the experiment, each target and distractor was presented an equal number of times and was presented equally in each vertical third of the screen (left, middle, right), ensuring that target and distractor were never presented in the same location in the same trial.

3.1.1.3 Results

Incorrect button presses and none responses were removed from the analysis included in the accuracy analysis. Median RTs for each of the eight conditions were calculated for each participant. The means of these medians are presented in Figure 16. along with the mean percentage of errors made in each condition.

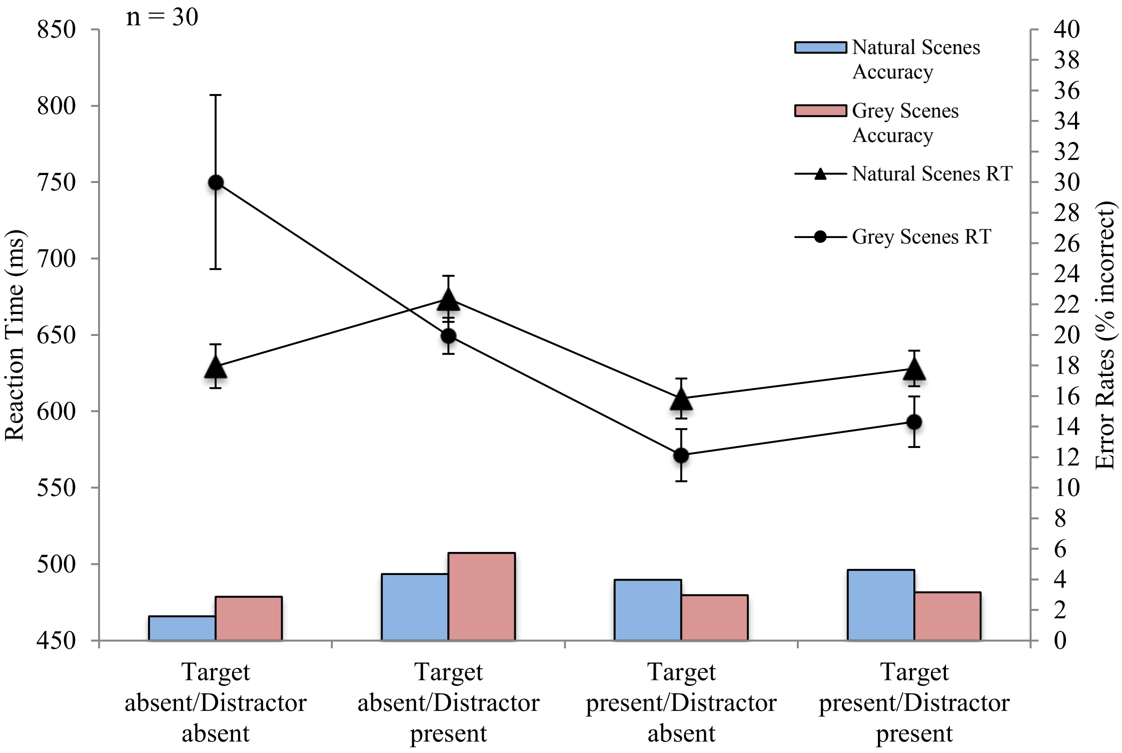


Figure 16. Means of the median correct detection rates (left axis) and error rates (right axis) for each condition of Experiment 4 for the target present conditions. The vertical error bars represent 95% within subject confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 (target: present, absent) × 2 (distractor: present, absent) × 2 (background: natural, grey) within-subjects ANOVA was conducted on the median RT data. This analysis yielded a significant main effect of target, $F(1, 29) = 30.53$, $MSE = 341448.98$, $p < 0.01$, $\eta^2 = 0.15$, suggesting that participants were faster to respond when a target was present ($M = 600.23$) compared to when a target was absent ($M = 675.66$). There was no significant main effect of distractor, $F(1, 29) = 0.13$, $MSE = 853.15$, $p > 0.05$, $\eta^2 = 0.00$, or significant main effect of background, $F(1, 29) = .20$, $MSE = 2211.30$, $p > 0.05$, $\eta^2 = 0.00$. The was a significant interaction between target × distractor, $F(1, 29) = 4.89$, $MSE = 35831.48$, $p < 0.05$, $\eta^2 = 0.02$. There was also a significant interaction between target × background, $F(1, 29) = 13.34$, $MSE = 106197.30$, $p < 0.01$, $\eta^2 = 0.05$, and a significant interaction between distractor × background, $F(1, 29) = 10.22$, $MSE = 76130.63$, $p < 0.01$, $\eta^2 = 0.03$. Importantly, there was a significant interaction observed between the target × distractor × background, $F(1, 29) = 10.87$, $MSE = 80978.634$, $p < 0.01$, $\eta^2 = 0.03$.

To explore the interaction of target, distractor and background, post hoc corrected pairwise comparisons were used. These revealed that for the natural target present scenes, mean RTs were significantly slower in scenes that contained a face distractor (target-present/distractor-present; $M = 628$) than RTs for scenes that did not (target-present/distractor-absent; $M=608.45$; $p < 0.05$, $d = 0.44$). For the natural target absent scenes, mean RTs were also significantly slower when a face distractor was present in the scene (target-absent/distractor-present; $M=673.67$) compared to RTs for scenes where it was absent (target-absent/distractor-absent; $M = 629.52$; $p < 0.05$, $d = 1.14$).

For the grey scenes, corrected pairwise comparisons also revealed that in the target present scenes, mean RTs were significantly slower when the scene contained a face (target-present/distractor-present; $M=593.11$) compared to scenes in which no face was present (target-present/distractor-absent; $M = 571.33$, $d = 0.47$). For the target absent trials, RTs in scenes that contained a face (target-absent/distractor-present; $M = 649.45$) were significantly faster than RTs in scenes that did not contain a face (target-absent/distractor-absent; $M = 750.12$; $p < 0.05$, $d = 0.60$).

Corrected pairwise comparisons also revealed that for the target-absent/distractor absent scenes mean RTs in grey scenes ($M = 750.02$) were significantly slower than RTs in natural scenes ($M = 629.52$; $p < 0.05$, $d = 0.63$). There was no difference in RTs for target-absent/distractor-present scenes between grey ($M = 649.45$) and natural ($M = 673.67$; $p > 0.05$, $d = 0.35$) scenes. RTs for target-present/distractor absent scenes were significantly slower in natural ($M = 608.45$) compared to grey ($M = 571.33$; $p < 0.05$, $d = 0.18$) scenes. Furthermore, RTs for target-present/distractor-present scenes were significantly slower in natural ($M = 628.00$) compared to grey scenes ($M = 593.12$; $p <$

0.05, $d = 0.60$). All other comparisons were not relevant to the hypotheses and are therefore reported in Appendix 2.

An analogous analysis on the error data revealed that there was no significant main effect of target, $F(1, 29) = 0.01$, $MSE = 0.13$, $p > 0.05$, $\eta^2 = 0.00$. There was a main effect of distractor, $F(1, 29) = 10.60$, $MSE = 157.536$, $p < 0.01$, $\eta^2 = 0.05$, suggesting that there were significantly more errors when the distractor was present ($M = 4.47\%$) relative to when the distractor was absent ($M = 2.85\%$).

There was no significant effect of background, $F(1, 29) = 0.01$, $MSE = 0.129$, $p > 0.05$, $\eta^2 = 0.00$. There was no significant target × distractor interaction $F(1, 29) = 4.06$, $MSE = 86.93$, $p > 0.05$, $\eta^2 = 0.03$. There was also a significant target × background interaction, $F(1, 29) = 7.00$, $MSE = 100.82$, $p < 0.05$, $\eta^2 = 0.03$. There was no significant distractor × background interaction, $F(1, 29) = 0.05$, $MSE = 0.51$, $p > 0.05$, $\eta^2 = 0.00$. There was also no target × distractor × background interaction, $F(1, 29) = 0.11$, $MSE = 1.16$, $p > 0.05$, $\eta^2 = 0.00$.

To examine the interaction between target and background paired comparisons were used, these revealed significantly more errors in grey target present scenes ($M = 4.31\%$) compared to target absent scenes ($M = 2.96\%$, $p < 0.05$, $d = 0.42$). No other comparisons were significant (all $p < 0.05$).

3.1.1.4 Discussion

Experiment 4 aimed to explore whether human faces are capable of capturing attention under conditions where the face was unrelated to the goal of the experimental task. The data suggests that in both the natural and grey scenes, the time taken to detect the ladybird target was slower when an irrelevant face also appeared in the scene compared with when the face was absent. This is consistent with expectations and can be explained in terms of attention capture: because the irrelevant face captures attention in a stimulus-driven bottom-up manner, it required a longer time before response could be initiated (Theeuwes, 1991, 1992, 1994, 1996). More specifically, attention is allocated to the face first and then, on average, half of the scene must be searched before the ladybird target is located. However, when the face is absent, only half of the scene, on average, must be searched until the target is located, explaining the shorter search times for the distractor face absent scenes. These findings support the notion that the detection advantages for the human face-only targets in Experiment 2 were the result of an inherent attentional advantage for faces, rather than some other factor. They are therefore also consistent with the hypothesis that the human face automatically captures a viewer's attention, which adds support to previous research suggesting that attention is captured by faces (e.g. Lewis & Edmonds, 2003, 2005, Langton et al., 2008; Ro et al., 2007; Theeuwes & van der Stigchel, 2006).

Further, the finding of a face distractor effect in the grey backgrounds suggests that throughout varying contexts, similar to that used by previous research, attention capture by faces is a robust and consistent effect. In addition, that the advantages for faces in Experiment 2 and the current experiment are unlikely to have been caused by an experimental artefact resulting from the use of natural backgrounds.

Given that the targets and distractors were highly salient when presented in the grey scenes, grey scenes were a measure of how low-level factors might be facilitating detection. Considering this, the data produced an interesting finding. It was observed that search in grey scenes containing a target was significantly faster than search in the natural scenes containing a target. However, there was no difference in search between natural and grey scenes that contained only a distractor (target-absent/distractor-present). This finding might suggest that some low-level factor of the target, but not the distractors, might be facilitating the faster responses in the grey scenes than the natural scenes. In turn, this could also indicate that low-level factors are also enabling the faster detection of the ladybird targets in the natural scenes. However, even if this were correct and some low-level factor of the targets were facilitating a faster response to the targets, the presence of a human face still slowed down search for the target. Therefore, the face distractor effect is robust enough to override any detection advantages of the target owing to low-level confounds.

The data produced another interesting finding worthy of discussion. In the natural scenes, for the target absent trials, the time taken to reject the presence of the target was increased by the presence of a distractor face, compared to when it was not present in the scene. The presence of the face distractor effect in the target absent trials suggests that once attention has been allocated to a face, whether this was preferentially attended or not, there is a particular difficultly in then shifting attention away from the face to search the remainder of the scene. Indeed, these results converge with an existing body of research indicating that it is particularly difficult to disengage attention from faces, as they may retain attention more than other objects (Bindemann, Burton, Hoode, Jenkins & de Haan, 2005c; Ro et al., 2007). In attention research, the distinction between attention capture and a delay in the ability to disengage attention once it has been allocated to a specific item has often been made (Langton et al., 2008). It therefore seems likely that the distractor face effect observed in Experiment 4 can be attributed to both attention capture by faces and also a general attentional bias that makes them hard to disengage processing resources away from.

On the other hand, in the grey target absent scenes, a similar face distractor effect was not observed. Instead, search times in scenes that contained the face distractor in isolation (target-absent/distractor-present) were significantly faster than search in the distractor-absent scenes, i.e., a blank grey screen (target-absent/distractor-absent trials). Given that in the natural target absent scenes faces appeared to retain attention, the opposite finding in the grey scenes might at first appear counterintuitive to the notion of an attention retention by faces. However, when the distractor face was present in the target-absent grey scenes, it was the only object present on the grey scene. The Guided Search (e.g. Wolfe, 1994) model of visual search proposes that attention is guided towards items in a scene in an order that is determined by a pre-attentive analysis. This is with regards to a map of saliency in which the most salient element has the highest activation and is attended to first. The attended item is then compared with the items in memory and if this item is not the target, this location is inhibited and the item of next most salience is then attended (see also Itti & Koch, 2000; Sagi & Julesz, 1985; Nothdurft, 2000). Taking this into consideration, any singular object in the grey scene, regardless of whether the object was a target or distractor, will be the only active area on the saliency map. Attention will therefore be automatically drawn to the location of the face, upon determining that this item is not the target and that no other active salient areas on the saliency map, search will terminate and a target absent response initiated. However, when the scene is both target and distractor absent, although there are no items to search, the whole scene must be examined before a target absent response can be initiated. Taken together, these factors might explain the longer latencies for the target-absent/distractor-absent scenes. Moreover, it is also possible that the slower responses in target-absent/distractor-absent scenes may reflect participants being unsure of the nature of the task, given that they were just presented with a grey scene.

3.1.2 Experiment 5

3.1.2.1 Introduction

In order to examine whether human bodies have an attentional advantage, Experiment 4 was replicated using headless human bodies in place of the body-less faces. It was hypothesized that if bodies capture attention, the presence of a task irrelevant body in the scene will slow search time for the ladybird target relative to when the body is not present. The observation of the body distractor effect would provide evidence in favour of the advantages observed for bodies in Experiment 2 as being attentional. Furthermore, the presence of a body distractor effect in the grey scenes will rule out the possibility that advantages for bodies are some artefact of presenting stimuli within natural backgrounds. Not only that but it also provides evidence for the body distractor effect being a consistent and robust effect in variable contexts.

3.1.2.2 Method

3.1.2.2.1 Participants

A convenience sample of 42 participants (38F, 4M) between 19 – 41 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 30 minutes.

3.1.2.2.2 Design, Stimuli, Apparatus and Materials, and Procedure

The design, stimuli, apparatus and materials, and procedure were all identical to Experiment 4. The only changes were that the distractor stimuli were the body stimuli as used in Experiments 1 – 4. Examples of the experimental conditions are shown in Figure 17.

Natural target-absent/ distractor-absent

Natural target-absent/ distractor-present

Natural target-present/ distractor-absent

Natural target-present/ distractor-present

Grey target-absent/ distractor-absent

Grey target-absent/ distractor-present

Grey target-present/ distractor-absent

Grey target-present/ distractor-present

Figure 17 An illustration of the stimuli for each of the experimental conditions in each phase of Experiment 5.

3.1.2.3 Results

Again, incorrect button presses and none responses were counted as errors and were removed before analysis and included in the accuracy analysis. Median RTs for each of the eight conditions were calculated for each participant. Figure 18. shows the means of the median correct RTs for all conditions along with the mean percentage of errors made in each condition.

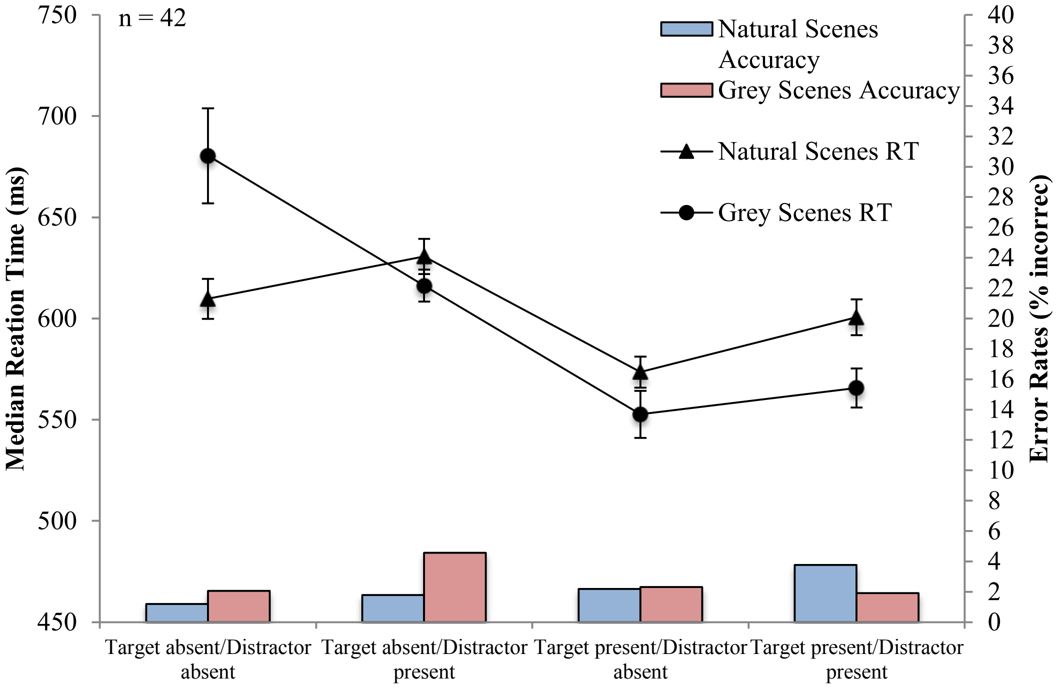


Figure 18. Means of the median correct detection rates (left axis) and error rates (right axis) for each condition of Experiment 5 for the target present conditions. The vertical error bars represent 95% within subject confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 (target: present, absent) × 2 (distractor: present, absent) × 2 (background: natural, grey) within-subjects ANOVA was conducted on the median RT data. This analysis yielded a significant main effect of target, $F(1, 41) = 46.42$, $MSE = 313785.19$, $p < 0.01$, $\eta^2 = 0.21$, suggesting that participants were faster to respond when a target was present (M=573.13) compared to when a target was absent (M=634.25). There was no significant main effect of distractor, $F(1, 41) = 0.03$, $MSE = 50.30$, $p > 0.05$, $\eta^2 = 0.00$, or significant main effect of background, $F(1, 41) = 0.00$, $MSE = 1.07$, $p > 0.05$, $\eta^2 = 0.00$.

There was a significant interaction between target × distractor, $F(1, 41) = 15.76$, $MSE = 36562.57$, $p < 0.01$, $\eta^2 = 0.02$. The was also a significant interaction between target × background, $F(1, 41) = 25.95$, $MSE = 65912.01$, $p < 0.01$, $\eta^2 = 0.04$, and a significant interaction between distractor × background, $F(1, 41) = 26.317$, $MSE = 51554.30$, $p < 0.01$, $\eta^2 = 0.03$. Importantly, there was a significant interaction between the target × distractor × background, $F(1, 41) = 11.68$, $MSE = 26447.50$, $p < 0.01$, $\eta^2 = 0.02$.

To explore the interaction with between target, distractor and background, post hoc corrected pairwise comparisons were used. These revealed that for the natural target present scenes mean RTs were significantly slower for scenes containing a body (target-present/distractor-present; $M = 600.64$) than RTs for scenes that did not contain a body (target-present/distractor-absent; $M = 573.52$; $p < 0.05$, $d = 0.74$). Further, for the natural target absent scenes mean RTs for scenes that contained a body (target-absent/distractor-present; $M = 630.63$) were significantly slower than mean RTs for scenes that did not contain a body (target-absent/distractor-absent; $M = 609.75$; $p < 0.05$, $d = 0.42$).

For the grey target present scenes, corrected pairwise comparisons confirmed that for mean RTs were significantly slower for scenes containing a body (target-present/distractor-present; $M=565.71$) than RTs for scenes that did not contain a body (target-present/distractor-absent; $M=552.65$; $p < 0.05$, $d = 0.50$). For grey target absent trials, RTs were significantly faster when a body was present (target-absent/distractor-present $M=616.24$) compared to scenes where no body was present in the scene (target-absent/distractor-absent; $M=680.39$; $p < 0.05$, $d = 0.68$).

Post hoc corrected pairwise comparisons also revealed that RTs in target-absent/distractor-absent were significantly faster for natural ($M = 609.75$) than grey scenes ($M = 680.39$; $p < 0.05$, $d = 0.57$). There was no difference in RTs between grey ($M = 616.24$) and natural target-absent/distractor-present scenes ($M = 630.63$; $p > 0.05$, $d = 0.79$). In addition, RTs for target-present/distractor-absent scenes were significantly slower for natural ($M = 573.52$) than grey scenes ($M = 552.65$; $p < 0.05$, $d = 0.88$). RTs for target-present/distractor-present scenes were significantly slower in natural scenes ($M = 600.64$) than in grey scenes ($M = 565.71$; $p < 0.05$, $d = 0.74$). All other comparisons were not relevant to the hypotheses and are therefore reported in Appendix 3.

An analogous analysis was conducted on the mean number of errors. There was no significant main effect of target, $F(1, 41) = 0.15$, $MSE = 1.86$, $p > 0.05$, $\eta^2 = 0.00$. This revealed a main effect of distractor, $F(1, 41) = 8.66$, $MSE = 97.03$, $p < 0.01$, $\eta^2 = 0.03$, with more errors being made in when a distractor was present (3.01%) compared to when absent (1.94%). There was no significant main effect of background, $F(1, 41) = 3.41$, $MSE = 19.32$, $p > 0.05$, $\eta^2 = 0.01$. There was no significant target × distractor interaction, $F(1, 41) = 1.54$, $MSE = 19.31$, $p > 0.05$, $\eta^2 = 0.01$. There was a significant target × background interaction, $F(1, 41) = 33.22$, $MSE = 150.67$, $p < 0.05$, $\eta^2 = 0.05$. There was no distractor × background interaction, $F(1, 41) = 0.00$, $MSE = 0.02$, $p > 0.05$, $\eta^2 = 0.00$. Importantly, there was a significant interaction observed between the target × distractor × background, $F(1, 41) = 7.69$, $MSE = 79.94$, $p < 0.01$, $\eta^2 = 0.03$.

Post hoc corrected pairwise comparisons were used to explore the interaction between target, distractor and background. These revealed that for the natural target present scenes, significantly more errors were made when a distractor was also present (target-present/distractor-present; M =3.77%) than when absent (target-present/distractor-absent; 2.18%; $p < 0.05$, $d = 0.43$). For the target absent scenes, there was no difference in the number of errors made when the distractor was present (target-absent/distractor-present; M =1.79%) than when absent (target-absent/distractor-absent; M =1.19%; $p > 0.05$, $d = 0.19$).

For the grey scenes, for the target present scenes, there was no difference in the number of errors made between the distractor present (target-present/distractor-present; M =1.92%) and distractor absent (target-present/distractor-absent; M =2.31%; $p > 0.05$, $d = 0.12$) scenes. For the target absent scenes, there were significantly more errors made when the distractor was present (target-absent/distractor-present; M =4.46%) than when the distractor was absent (target-absent/distractor absent; M = 2.05%; $p < 0.05$, $d = 0.42$).

Furthermore, there were no significant differences in the number of errors made between the natural (M = 1.19%) and grey target-absent/distractor absent scenes (M = 2.05%, $p > 0.05$, $d = 0.16$). In the target-absent/distractor-present scenes, there were significantly more errors in the grey scenes (M = 4.5%) than the natural scenes (M = 1.79%, $p < 0.05$, $d = 0.77$). In the target-present/distractor-absent scenes, there was no difference in the number of errors made between the natural (M = 2.18%) and grey (M = 2.31%) scenes ($p > 0.05$, $d = 0.04$). For target-present/distractor-present scenes there were significantly more errors made in natural scenes (M = 3.77%) than grey scenes (M = 1.91%, $p < 0.05$, $d = 0.53$). All other comparisons were not relevant to the hypotheses and are therefore reported in Appendix 4.

3.1.2.4 Discussion

The aim of Experiment 5 was to expand on the findings of Experiment 4 by examining whether human bodies, like faces, also capture attention. The data from Experiment 5 suggests that in both phases the time taken to detect a ladybird target increased in the presence of the task-irrelevant human body. This body distractor effect, like the face distractor effect, can be explained in terms of attention capture. Specifically, because the body captures attention in a stimulus-driven bottom-up manner, attention is oriented to the body first in the scene before the rest of the scene can be searched, resulting in longer search times than when the body is not present (Theeuwes, 1991, 1992, 1994, 1996). In line previous research (Ro et al., 2007; Downing et al., 2004) the results from Experiment 5 suggest that human bodies may capture attention. On that basis, like faces, it therefore appears that the detection advantages for bodies observed in Experiment 2 are attentional. The presence of the body distractor effect in the grey backgrounds also suggests that the attentional biases for bodies in the current research are unlikely to have been caused by experimental artefact of the natural backgrounds. It also suggests that the effect is robust and consistent across various variable-viewing conditions.

Interestingly, for the target absent trials in the natural scenes, the time taken to reject the presence of the target was increased by the presence of a distractor body, compared to when the body was not present in the scene. Similar to the results found for faces, the most conceivable explanation of this result is that once attention has been allocated to a body, the body distractors retain a viewer’s attention, making it difficult for participants to disengage and search the remainder of the visual scene. The findings of this experiment concur with previous findings that attention was hard to disengage from threat-related body postures (Bannerman, Milders, & Saharie, 2010), albeit here we used neutral body postures. Again, similar to the distractor face effect, it seems likely that the distractor body effect observed in the current experiment is attributable to a combination of both attention capture and retention by the presence of the bodies.

Again, in the grey target absent scenes, search was faster when the distractor was present compared with when it was absent. As for lack of a face distractor effect in Experiment 4, the lack of a distractor body effect can be explained in terms of participants being unsure of the nature of the task when merely presented with a blank grey screen in the experiment. This is supported by search being significantly slower and more error prone in the grey target-absent/distractor-absent scenes, than the equivalent natural scenes.

In addition, the data produced another interesting finding. Mirroring the results from Experiment 4, it was observed that search in scenes that contained a target were significantly faster for grey scenes than natural scenes. There was however no difference in search between the natural and grey scenes that contained only a body (target-absent/distractor-present). In line with the conclusions from Experiment 4, these findings suggest that some low-level image based property of the ladybird targets was facilitating faster responses to these in the grey scenes. However, despite this, the body distractor slowed the search for the target.

3.1.3 Experiment 6

3.1.3.1 Introduction

The findings of Experiments 4 and 5 appear to be consistent with evidence for attention capture by human faces and bodies (Langton et al. 2008; Ro et al., 2007; Theeuwes & van der Stigchel, 2006; Downing et al., 2004; Bindemann et al., 2010). However, it is well-established from research using the traditional visual search paradigm that an increase in total number of elements in a display (i.e. set-size) produces a robust degradation of performance in visual search response and accuracy, consistent with the serial search of selective attention (Bundesen 1990; Bundesen, Habekost & Kyllingsbaek, 2005; Desimone & Duncan 1995; Duncan & Humphreys 1989; Treisman & Gelade 1980; Wolfe, 2007). Considering this, it would be expected that search in a scene with two items should take longer than search in a scene with one item. Therefore, the face and body distractor effect observed in Experiments 4 and 5, respectively, may reflect the time costs of serial search taking longer to search scenes that contained two items, target and distractor (target-present/distractor-present), compared to scenes that only contained the target in isolation (target-present/distractor-absent). To address this issue the experiment was repeated so that all target and distractor present scenes contained two objects.

It was hypothesized that if bodies capture attention the search for the ladybird targets should be slowed by the presence of an irrelevant body in the display as compared to when the body is absent but two target ladybirds are present. However, if the time costs of serial search is responsible for the body distractor effect, and by analogy the face distractor effect, then there should be no difference between scenes that contain two targets and scenes that contain both a target and distractor. The experiment was completed using both natural and grey backgrounds. Again, the grey backgrounds were a measure of how search is influenced by low-level confounds.

It should, however, be noted that the presence of two ladybird targets in a scene doubles the probability of target detection. A consequence of this could be that search in the target-target condition may be faster than search in the target-distractor condition, simply because there is more target information in the scene. As such, the body distractor effect may appear in Experiment 6 not because the body distractors slow responses to the target, but as a consequence of faster responses in the target-target condition. In order to examine this possibility additional analyses were conducted. It was therefore hypothesized that if a body distractor effect in Experiment 6 is facilitated by faster target search owing to the presence of two targets in the target-target scenes (previously target-present/distractor-absent), then search in scenes containing two targets should be faster than search in scenes containing one target. However, to address this hypothesis, a comparison between the response times for natural target-target scenes (Experiment 6) and target-present/distractor-absent condition (Experiment 5) was required. However, given that two different sets of participants were used across the two experiments in order to make this comparison it was first necessary to ensure that these participants were responding in a similar way. To achieve this, the RTs for the natural target-distractor scenes from Experiment 6 were compared to RTs for the natural target-present/distractor-present scenes from Experiment 5. Given this is the only condition that is identical in both experiments, it was reasoned that if performance on these conditions was similar, it could be tentatively suggested that participants were responding in a similar way in all conditions across both experiments. Following this, a comparison could then be made between the RTs from target-target scenes (Experiment 6) and target-present/distractor-absent scenes (Experiment 5).

3.1.3.2 Method

3.1.3.2.1 Participants

A convenience sample of 30 participants (25F, 5M) between 19 – 24 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 25 minutes.

3.1.3.2.2 Design, Stimuli, Apparatus and Materials, and Procedure

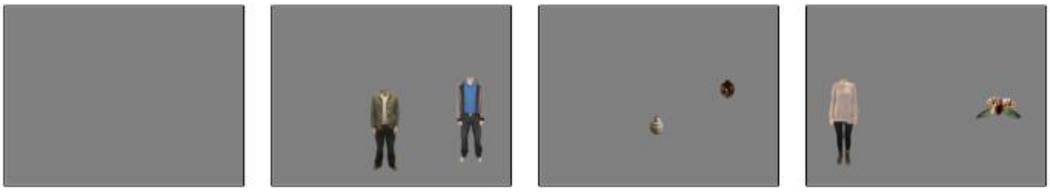
The design, stimuli, apparatus, materials, and procedure were all identical to those of Experiment 5. The only difference was that now all of the target or distractor present scenes contained two objects. These four conditions were: i) two targets in a scene (target-target condition); ii) two distractors in a scene (distractor-distractor condition); iii) one target and one distractor in a scene (target-distractor condition); and iv) a scene with neither target nor distractor (target-absent/distractor-absent condition). Examples of the stimuli and experimental conditions are shown in Figure 19.

Natural target-absent/ distractor-absent

Natural distractor-distractor

Natural target-target

Natural target-distractor



Grey target-absent Distractor-absent

Grey distractor-distractor

Grey target-target

Grey target-distractor

Figure 19. An illustration of the stimuli for each of the experimental conditions in each phase of Experiment 6.

3.1.3.3 Results

As in previous experiments, incorrect button presses and none responses were counted as errors and were removed before analysis of the RT data. For each of the conditions median RTs for each participant was calculated. The means of these medians are displayed in Figure 20.

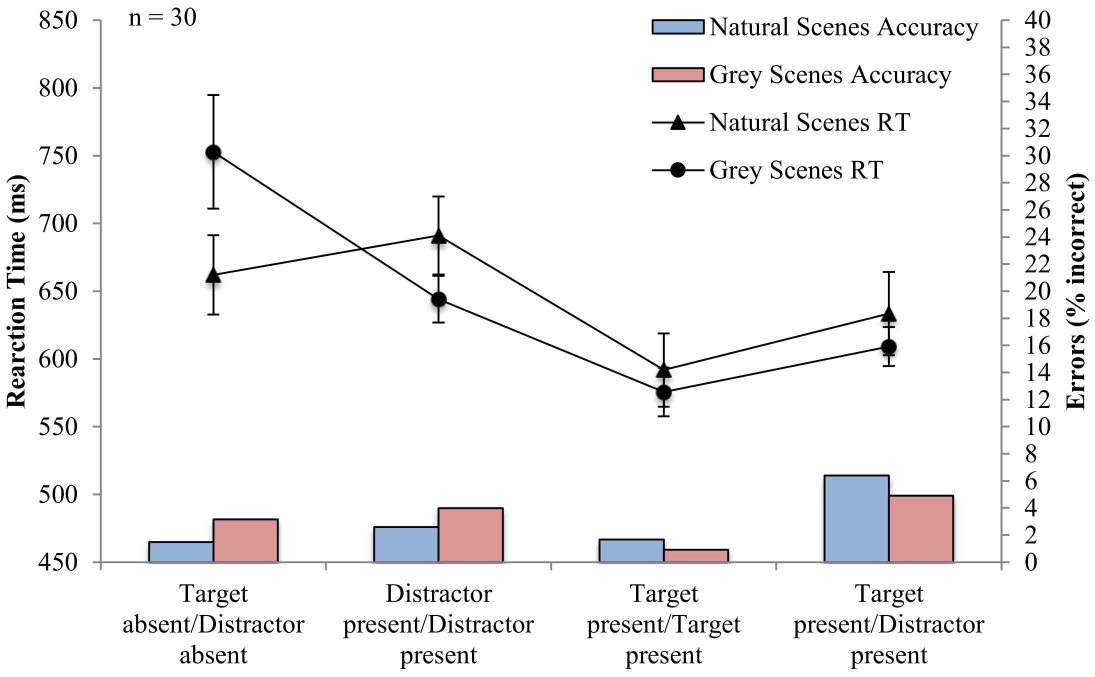


Figure 20. Means of the median correct detection rates (left axis) and error rates (right axis) for each of the experimental conditions in Experiment 6. Vertical bars represent the within subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 (background: natural, grey) × 4 (target/distractor arrangement: target-absent/distractor-absent, distractor-distractor, target-target, target-distractor) × 2 (background: natural, grey) within-subjects ANOVA was conducted on the median RT data. This revealed that there was no significant main effect of background, $F(1, 29) = 0.00$, $MSE = 34.50$, $p > 0.05$, $\eta^2 = 0.00$. There was however a significant main effect of target/distractor arrangement $F(3, 87) = 14.71$, $MSE = 523462.96$, $p < 0.01$, $\eta^2 = 0.16$). There was also a significant interaction between background × target/distractor arrangement, $F(3, 87) = 5.97$, $MSE = 169580.471$, $p < 0.01$, $\eta^2 = 0.05$.

To examine the interaction between background and target/distractor arrangement, post hoc corrected pairwise comparisons were conducted. These revealed that for the natural target present scenes, RTs were significantly slower when a distractor body was present (target-distractor; $M = 633.50$) compared to when two targets were present in the scene (target-target; $M = 591.83$; $p < 0.05$, $d = 0.81$). For the target absent scenes, RTs were significantly slower when two distractors were present in a scene (distractor-distractor; $M = 691.10$) compared to when no distractor was present in the scene (target-absent/distractor-absent; $M = 662.02$; $p < 0.05$, $d = 0.45$).

For the grey target present scenes, corrected pairwise comparisons revealed that RTs were significantly slower when a distractor was also present (target-distractor; $M = 609.13$) compared to when two targets were present in the scene (target-target; $M = 575.52$; $p < 0.05$, $d = 0.88$). In the target absent scenes, RTs were significantly faster when two distractors were present in a scene (distractor-distractor; $M = 644.07$) compared to when no distractors were present (target-absent/distractor-absent; $M = 752.77$; $p < 0.05$, $d = 0.71$). There was no difference between any of the natural scenes and their analogous grey counterpart scenes (all $p > 0.05$). All other comparisons were not relevant to the hypotheses and are therefore reported in Appendix 5.

An analogous analysis was conducted on the error data. This revealed that there was no significant main effect of background $F(1, 29) = 1.07$, $MSE = 27.04$, $p > 0.05$, $\eta^2 = 0.01$. There was a significant main effect of target/distractor arrangement $F(3, 87) = 8.70$, $MSE = 431.68$, $p < 0.01$, $\eta^2 = 0.11$. There was also a significant interaction between background × target/distractor arrangement $F(3, 87) = 7.40$, $MSE = 276.59$, $p < 0.01$, $\eta^2 = 0.07$.

To examine the interaction between background and target/distractor arrangement post hoc corrected pairwise comparisons were conducted. These revealed that for the natural scenes there were significantly more errors when both a target and distractor were present (target-distractor, $M = 6.39\%$) than when two targets were present in the scenes (target-target, $M = 1.67\%$, $p < 0.01$, $d = -0.722$).

For the grey scenes there were significantly more errors when both a target and distractor were present (target-distractor, $M = 4.91\%$) than when two targets were present in the scenes (target-target, M = 0.93%, p < 0.01, d = -0.915). All other comparisons were not relevant to the hypotheses and are therefore reported in Appendix 6.

Additional analyses were conducted to examine whether participants were responding faster to the target-target condition, because the presence of two targets in a scene doubled the probability of target detection. An independent samples t-test revealed that there was no significant difference between the RTs from target-present/distractor-present scenes in Experiment 5 (M = 583.16) and RTs from the Target-Distractor scenes in Experiment 6 (M =621.32), t(70) = -1.54, p > 0.05, d = 0.36. This appears to suggest that participants were responding similar in terms of their RTs in both Experiments 5 and 6. Considering this, it was felt that a comparison between the RTs for the target-present/distractor-absent condition (Experiment 5) and target-target scenes (Experiment 6) could be made. Consequently, an independent samples t-test revealed that there was no significant difference between the RTs from target-present/distractor-absent scenes in Experiment 5 (M = 563.09) and RTs from target-target scenes in Experiment 6 (M = 583.68), t(70) = 0.78, p > 0.05, d = 0.22.

3.1.3.4 Discussion

Experiment 6 aimed to examine whether the face and body distractor effect observed in Experiments 4 and 5 were the result of the time costs of serial search when examining both target and distractor (target-present/distractor-present trials) compared with search examining one target (target-present/distractor-absent trials). To investigate this possibility, neither the target nor distractor was ever presented in a scene in isolation; they were always simultaneously presented with another target or distractor. Consistent with the findings from Experiments 4 and 5, the results of Experiment 6 demonstrate that, irrespective of background, the time taken to detect the presence of a ladybird target was increased by the presence of an irrelevant body in the scene, compared with scenes that contained two ladybird targets. Therefore, given that in both the target-target and target-distractor scenes there were an equal number of items to search, the results provide supporting evidence that the body distractor effect in Experiment 5, and by extension the face distractor effect in Experiment 4, were the result of attention capture by bodies and faces, respectively.

However, it was suggested that the body distractor effect in Experiment 6 could be facilitated by faster target search in the target-target trials because of the presence of two targets, rather than by the presence of a body distractor slowing the search for the target. A comparison between the search times for target-present/distractor-absent condition (Experiment 5) and target-target scenes (Experiment 6) revealed no difference in search for the target. Therefore, it can be tentatively suggested that the presence of two targets in a scene does not appear to facilitate faster search relative to search where only one target is present.

3.1.4 General Discussion

The experiments reported in this chapter used a variation of the ASP to examine whether human faces and bodies capture attention in a stimulus-driven bottom-up manner. As discussed in the introduction, the ASP is a more stringent test of attention than the visual search paradigm used in Chapter 2 (Experiments 1, 2, 3). The data from Chapter 3 have revealed a number of interesting findings. The key finding being that regardless of the background RTs to the ladybird target were slower when an irrelevant face or body was present, than if (face or body) was absent. In line with previous findings for faces (e.g. Bindemann et al., 2005; Hershler & Hochstein, 2005; Theeuwes & Van der Stigchel, 2006; Ro et al., 2007), and bodies (Downing et al., 2004; Ro et al., 2007; Bindemann et al., 2010), the results presented here support the notion that faces and bodies capture attention, ensuring that attention is always orientated towards them first, when they appear in a scene. Since all of the results were obtained using the ASP it would appear that the effect is a bottom-up stimulus-driven effect.

Also of interest was the finding that when presented in target absent natural backgrounds, the presence of a distractor (face or body) produced a delay on search compared to when absent. This distractor effect in the target absent scenes suggests that faces and bodies may also be particularly efficient at retaining visual attention, such that it is difficult for participants to disengage attention and search the remainder of the visual scene. This finding is consistent with existing research suggesting that both faces (Bindemann et al., 2005; Ro et al., 2007) and bodies (Bannerman et al., 2010; Ro et al., 2007, experiment 4) are able to retain visual attention for longer than other objects. Consequently, the results in the current chapter strongly suggest that the face and body distractor effect arise because they both capture attention and because they then retain attention, making it hard to disengage processing resources from them.

Another finding worthy of discussion was that in Experiment 6 there was no difference in search times between any of the natural scenes and their comparable grey scene counterparts. Conversely, in Experiments 4 and 5, search had been faster in the grey scenes that contained a target compared to natural scenes that contained a target. This indicated that some low-level factor of the target might be facilitating their detection. The failure to replicate this result in Experiment 6 suggests that the low-level properties of the targets no longer facilitate faster detection. Rather, these results may instead reflect a difference in the visual perceptual load. Note, although there is no definitive rule about what defines perceptual load, a greater number of items within a scene has been taken to suggest a higher perceptual load (Lavie, Lin, Zokaei & Thoma, 2009). Perceptual load theory (Lavie, 1995) postulates that processing will occur automatically until processing capacity is full. It also suggests that under conditions of low-perceptual load any spare processing after the relevant stimuli will unintentionally spill over to task-irrelevant stimuli. Conversely, under high perceptual load, task-relevant information exhausts resources (early selection) and task-irrelevant stimuli will not be processed.

Therefore, in the grey target-present/distractor-absent scenes of Experiments 4 and 5, each scene only contained one target, hence the visual load was low and since there was nothing else to process in the scene, search was relatively fast. Conversely, although only one target was present in the natural target-present/distractor-absent trials, they were presented in natural backgrounds; hence the perceptual load was higher. Specifically, although the backgrounds might be task-irrelevant, items in the background may share a number of task-relevant features with the target of search. Therefore, search for the target becomes increasingly more challenging, presumably because these areas will capture attentional resources, slowing search in the natural scenes (Wolfe, Oliva, Horowitz, Butcher & Bompa, 2002; Vavassis & Von Grünau, 2007). Conversely, in Experiment 6, two targets were presented in both the natural and grey target-target scenes; hence the task-relevant visual load is increased in both scene types. Therefore, in the grey target-target scenes there are two task-relevant stimuli that may slow processing. In the natural scenes, there were also two task-relevant stimuli, which may exhaust processing capacity, regardless of the shared target features in the background. Taken together, this interpretation could explain the equal detection between the natural and grey (target-target) scenes in Experiment 6. However, whilst this interpretation is appealing, target-present/distractor-present scenes and target-distractor scenes were identical across both experiments, in Experiments 4 and 5 responses in the grey scenes were faster. Conversely, in Experiment 6 there was no difference in search times. Since the visual load in these scenes in both Experiment 5 and 6 was identical, an increase in visual load cannot explain the equal detection rates in Experiment 6. This suggests that the difference in the results between the target-present/distractor-absent and target-target scenes may not be owing to a difference in visual load, but instead reflects some artefact of the experiment.

Finally, demonstrating the distractor effect (for both faces and bodies) irrespective of background validates the notion that advantages for human targets in Experiment 2 and attention capture by faces and bodies in the current chapter are not the result of some experimental artefact owing to the use of natural backgrounds. In turn, this suggests that previous findings of attention capture by faces and bodies may not necessarily be confined to uniform colour backgrounds and may transfer to richer more complex environments (faces: Bindemann & Burton, 2009; Bindemann, et al., 2007; Langton et al., 2008; Ro, Friggel, & Lavie, 2007; Ro et al., 2001; Theeuwes & Van der Stigchel, 2006; bodies: Ro et al., 2007; Downing et al., 2004; Mack & Rock, 1998). Therefore, attention capture by faces and bodies appears to be a robust effect. Given this, the rest of the thesis will present stimuli in natural backgrounds.

In sum, the irrelevant distractor face and body effect demonstrated in this chapter can be explained in terms of faces and bodies capturing attention and also because of a difficultly in an ability to disengage attention from faces and bodies. There are, however, a number of shortcomings to these conclusions, which will be explored in Chapter 4.

Chapter 4

4 Non-Body Distractors in The Additional Singleton Paradigm

4.1 Experiment 7

4.1.1 Introduction

Chapters 2 and 3 examined attention capture using human faces and bodies as visual stimuli. Although much of the face research has used artificial displays to present face stimuli, the experiments in this thesis have established that attention capture by faces appears to be a robust effect regardless of the background it is presented in (see Experiment 4). Therefore, given that the results from previous experiments are consistent with research indicating that faces appear to capture and retain attention (e.g., Bindemann & Burton, 2008; Langton et al., 2008; Ro et al., 2007; Theeuwes & Van der Stigchel, 2006; Bindemann et al., 2005; Hershler & Hochstein, 2005; Mack et al., 2002), the following experiments will deal exclusively with attention capture by human bodies.

Whilst the findings from Chapter 2 appear to be consistent with the notion of attention capture by bodies using the ASP, there are a number of potential experimental shortcomings that need addressing. First, although human bodies are viewed on a daily basis, they more often viewed in combination with the human face. Given this, observers may not have frequently perceived a body without a head and may be unfamiliar with headless human bodies. In line with this suggestion, research on visual search has shown that observers can rapidly detect an odd stimulus amongst an array of distractors (e.g. Triesman & Gelade, 1980). Furthermore, search for a familiar target in scenes containing unfamiliar distractors has been shown to be less efficient than search for an unfamiliar target in a scene of familiar distractors (Shen & Reingold, 2001; Wolfe, 2001; Wang, Cavanagh & Green, 1994). Specifically, Wolfe (2001) demonstrated that it was easier to search for a 'dead elephant' (inverted) surrounded by 'alive elephants' (upright) than it was to search for an alive elephant surrounded by dead elephants. Wang et al. (1994) also argued that familiar objects require less processing, whereas, unfamiliar objects require more processing. Hence, when the target and distractor in a scene are familiar, attention is not differentially draw to either.

Conversely, when either the target or distractor is unfamiliar, attention is drawn to the unfamiliar one, hence when a distractor is unfamiliar it causes a delay in the search for a target. As Wang et al. (1994) suggest, familiarity (or its inverse, unfamiliarity) appears to be a feature that will pop out in visual search. It is important to point out at this stage that, in referring to be stimuli as being unfamiliar, we mean unusual in a sense that they deviate from their normal appearance and are therefore, odd. For the remainder of this thesis the word odd will be used to denote unfamiliarity/unusualness. In turn, the human body-only targets (no heads) in Experiments 1 – 6 may have captured attention owing to them being odd stimuli. To address this potential problem, in Experiment 7 the human body distractors were presented as full-body silhouettes. Human body silhouettes are a highly familiar body representation and are perceived on a regular basis, e.g. as a body shadow. Therefore, they control the possibility that they are odd. Furthermore, although body silhouettes contain the human head, they do not contain any facial information, as a result any observation of attention capture by body silhouettes cannot be attributed to the face (Downing et al., 2004).

Importantly, neuroimaging research has demonstrated that the extrastriate body area (EBA: Downing et al., 2001), the region within the human brain that selectively responds to the visual representation of the human body and body parts, also responds higher to images of the full-human body silhouettes relative to control stimuli (Downing et al., 2001; Peelen and Downing, 2005; Schwarzlose et al., 2005; Spiridon et al., 2006; Weiner & Grill-Spector, 2010). One implication of this is that body silhouettes will elicit the same behavioural responses as a non-silhouette body. Consistent with this notion is that several behavioural studies have found evidence supporting a general attentional bias favouring body representations. For instance, when using the inattentional blindness paradigm, Mack and Rock (1998) found that critical stimulus detection rates are higher when the critical stimulus was stick figure of a person, rather than a schematic drawing of a tree or house. Though not conclusive, it would appear that basic body schematics seem to show preferential processing (c.f. attention capture) over other stimuli. However, importantly, Downing et al. (2004) observed that critical human body silhouette stimuli were more likely to be detected than critical non-body object targets (e.g. object silhouettes) in the inattentional blindness paradigm. They also found that critical human body silhouettes were more likely to be detected than either critical human hand silhouettes or scrambled versions of bodies, therefore ruling out detection advantages based on biological relevance or low-level image properties, respectively. As a consequence, Downing et al. (2004) concluded that bodies and body representations were given priority for attentional selection in comparison to the controls used in their experiments.

Taken together, the above findings suggest that in the visual system, there is a general preference for bodies across a range different exemplars and images format. In order words, the mechanism that facilitates body detection does not have a preference for non-silhouette or silhouette bodies. Therefore, it was hypothesized that if bodies capture attention then the presence of a body silhouette should slow search for the target relative to when it is absent. However, if the headless bodies in Experiments 5 and 6 captured attention because they were odd, then the presence of a body silhouette should not slow search for the target.

Second, a major limitation of Chapter 2 was that there was no non-body control distractor in Experiments 5 and 6. Although the presence of a task-irrelevant body was shown to slow search for a target, this distractor effect was not compared to the effect of a non–body distractor stimulus. Without this comparison, it is impossible to determine whether the distractor effect is unique to bodies. As discussed in Chapter 3, it is known that an increase in set-size leads to an increase in search times consistent with the serial search of selective attention (Bundesen 1990; Bundesen et al. 2005; Desimone & Duncan 1995; Duncan & Humphreys 1989; Treisman & Gelade 1980; Wolfe 2007). Therefore, it is expected that search in scenes containing two items will be slower than scenes containing one item to some degree – although this does not appear to be an issue if both items are targets (see Experiment 6). However, if human bodies capture attention, then their presence in a scene should slow search for a target more than other objects should slow search for a target. To examine this possibility, in Experiment 7, a non-body control was also used as a task-irrelevant distractor. For this purpose, lamp silhouettes were used. Lamps are highly familiar objects that are symmetrical and generally viewed in an upright position. It was predicted that if bodies capture attention, then search in scenes containing both a target and body distractor should be significantly slower than search in scenes containing a target and lamp distractor.

Finally, in Experiments 5 and 6, the distractor human body stimuli were always much larger in the scene than the target ladybirds. Previous research has shown that larger objects capture attention in a stimulus-driven bottom-up manner that is independent of top-down influences (Proulx & Egeth, 2008; Proulx, 2010). Therefore, it is possible that the distractor bodies captured attention in Experiments 5 and 6 because they were larger in size than the ladybird targets. Note that this interpretation cannot be applied to the faces in Experiment 4, as they were approximately the same size as the ladybird targets. Experiment 7 aimed to eliminate this possibility by using a target that was matched to size of the body distractors. For this purpose, silhouette grandfather clocks were used as the targets. It was predicted that if human bodies capture attention, search time in scenes containing a grandfather clock target and body distractor should be slower compared with search in scenes where no body distractor was present. However, if the relative size of the body distractors in comparison to the smaller ladybird targets captured attention in Experiments 5 and 6, then when matched for size the human bodies will not slow search for the grandfather clock target.

4.1.2 Method

4.1.2.1 Participants

A convenience sample of 41 participants (34F, 7M) aged between 18 – 25 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 40 minutes.

4.1.2.2 Design

A 2 × 3 within-subjects design was employed with two factors: target (present vs. absent) and distractor (absent, lamp-distractor present, body-distractor present). The dependent variables were the reaction time (milliseconds) taken from stimulus onset and participants accuracy. The experiment used an ASP paradigm and the data were analysed using a 2 × 3 analysis of variance. Post hoc analysis (where appropriate) used paired *t* tests with a false-discovery rate correction for multiple comparisons (Benjamini and Hochberg, 1995).

4.1.2.3 Stimuli

The 24 backgrounds, 12 full-body, and 12 target grandfather clock stimuli were identical to those from Experiments 1 – 6. The lamp distractor stimuli were created from 12 different photographic images of long-stand lamps from an open source on the internet. They were cropped out of their natural outline and manipulated to be approximately the same size (height) as the target grandfather clocks and body distractors. All targets and distractors were then transformed into black silhouettes by applying a black filter. All manipulations were completed in ADOBE Photoshop (CS 8).

For the target present scenes, the grandfather clock target was presented in isolation in the scene with no distractor present (target-present/distractor-absent condition), or the grandfather clock target was presented simultaneously with a body distractor in the scene (target-present/body-distractor condition), or the grandfather clock target was presented simultaneously with a lamp distractor in the scene (target-present/lamp-distractor condition).

For the target absent scenes, the scene contained neither the target nor distractor (target-absent/distractor-absent condition), or the human body distractor in isolation in the scene (target-absent/body-distractor condition), or the lamp distractor in isolation in the scene (target-absent/lamp-distractor condition). Example stimuli illustrating these experimental conditions can be seen in Figure 21.

Target-absent/distractor-absent

Target-absent/body-distractor present

Target-absent/lamp-distractor present

Target-present/distractor-absent

Target-present/body-distractor present

Target-present/lamp-distractor present

Figure 21. An illustration of the stimuli for each of the experimental conditions in Experiment 7.

Apparatus and Materials

The apparatus and materials for Experiment 7 were identical to that used in Experiments 1 – 6.

4.1.2.4 Procedure

The procedure for Experiment 7 was identical to the one used in Experiments 4 – 6. The participants received 216 trials, comprising of 108 target present scenes (36 trials for target-present/distractor-absent scenes, 36 trials for target-present/body-distractor present, 36 trials for target-present/lamp-distractor present), and 108 target-absent scenes (36 trials for target-absent/distractor-absent scenes, 36 trials for target-absent/body-distractor present scenes, 36 trials for target-absent/lamp-distractor present scenes). On each trial, Psychopy selected a background scene at random so that each background was used an equal number of times. In target present scenes, Psychopy also randomly selected a target. Similarly, in distractor present scene, a distractor was also selected at random. Over the course of the experiment, each target and distractor was presented an equal number of times and presented equally within each vertical third of the screen (left, middle, right), ensuring that on each trial target and distractor were never presented in the same location. The trials were randomized for each participant.

4.1.3 Results

As before, incorrect button presses and none responses were counted as errors and were removed before analysis of the RT data. Median RTs for each of the conditions were calculated for each participant. The means of these medians are presented in Figure 22. along with the mean percentage of errors made in each condition.

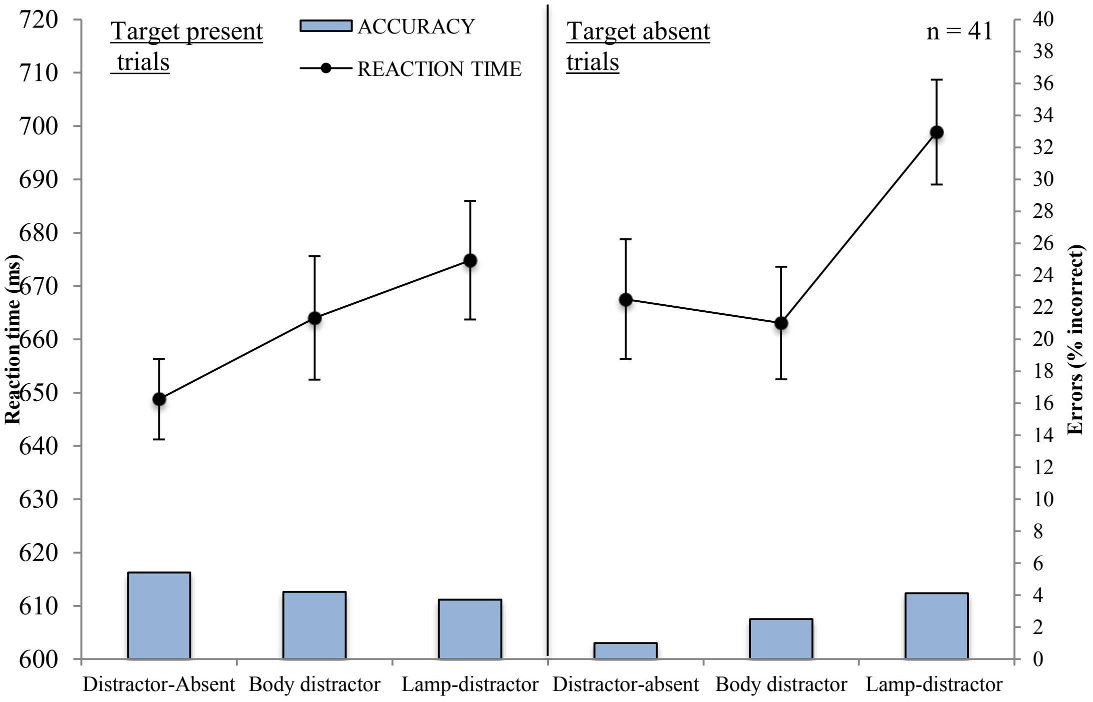


Figure 22. Means of the median correct detection rates and error rates for each condition of Experiment 7 for the target present and absent conditions. Vertical bars presented the within subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 (target: present, absent) × 3 (distractor: absent, body-distractor present, lamp-distractor present) within-subjects ANOVA was conducted on the median RT data. This analysis showed no significant main effect of target, $F(1, 40) = 2.57$, $MSE = 11970.146$, $p > 0.05$, $\eta^2 = 0.02$. There was a significant main effect of distractor, $F(2, 80) = 13.46$, $MSE = 19072.96$, $p < 0.01$, $\eta^2 = 0.08$.

Corrected pairwise comparisons revealed no significant difference in RTs between distractor-absent scenes (*M* = 658.15) and body-present scenes (*M* = 663.54, $p > 0.05$, $d = 0.16$). Mean RTs for lamp-distractor present scenes (*M* = 686.84) were significantly slower than mean RTs for body-present scenes (*M* = 663.54, $p < 0.05$, $d = 0.52$) and distractor-absent scenes (*M* = 658.15, $p < 0.05$, $d = 0.88$). There was no significant target × distractor interaction, $F(2, 40) = 2.13$, 3558.09, $p > 0.05$, $\eta^2 = 0.01$.

An analogous analysis was conducted on the error data, this revealed a main effect of target, $F(1, 40) = 12.56$, *MSE* = 221.32, $p < 0.01$, $\eta^2 = 0.08$, with significantly more errors for target present trials (*M* = 4.45%) as compared to target absent trials (*M* = 2.55%). There was no main effect of distractor, $F(2, 80) = 1.19$, *MSE* = 11.70, $p > 0.05$, $\eta^2 = 0.00$, but there was an target × distractor interaction, $F(2, 80) = 10.65$, *MSE* = 119.22, $p < 0.01$, $\eta^2 = 0.08$.

Pairwise comparisons were used to explore the target and distractor interaction. These revealed that there were significantly more errors in the target-absent/body-distractor present (*M* = 2.51%, $p < 0.05$, $d = 0.45$), target-absent/lamp-distractor present (*M* = 4.13%, $p < 0.05$, $d = 0.67$), target-present/distractor-absent (*M* = 5.42%, $p < 0.05$, $d = 0.89$), target-present/body-distractor present (*M* = 4.20%, $p < 0.05$, $d = 0.71$), and target-present/lamp-distractor present scenes (*M* = 3.73%, $p < 0.05$, $d = 0.78$) as compared to the target-absent/distractor-absent trials (*M* = 1.02%). Furthermore, there were significantly more errors for target-present/distractor-absent scenes (*M* = 5.42%) than for target-absent/body-distractor present trials (*M* = 2.51%; $p < 0.05$, $d = 0.58$) and for target-present/body-distractor present trials (*M* = 4.20, $p < 0.05$, $d = 0.34$). No other significant differences were observed, $p > 0.05$.

4.1.4 Discussion

Following some concerns over the use of headless bodies, the lack of a control distractor and target size, the aim of Experiment 7 was examine whether the body distractor effect, as observed in Experiments 5 and 6, was still present with these concerns addressed. It was hypothesized that if human bodies capture and retain attention more than other objects, then the body distractor effect should be larger than a distractor effect from the lamps (i.e., slower target search in the presence of a body compared to a lamp). However, the results showed that, regardless of whether the target was present or absent, search was significantly slower for scenes that contained a lamp distractor compared to scenes that contained a body distractor. The results also demonstrated that search was significantly slower in the scenes that contained a lamp distractor compared to search in scenes that did not contain a lamp. Interestingly, there was no difference in search between body-distractor present scenes and search in distractor absent scenes. Taken together, these results suggest the presence of a lamp distractor effect, but the absence of a similar effect for the body distractors.

The absence of the body distractor effect strongly suggests that body silhouettes and by extension bodies do not capture attention. If they did, their presence in a scene would have slowed search more than the presence of a distractor lamp. In fact, the body silhouettes do not slow search for the grandfather clock targets at all. These findings therefore fail to replicate the results of Experiment 5 and 6 and are also inconsistent with previous findings of attention capture by silhouette bodies and body parts (Downing et al., 2004; Ro et al., 2007).

There a number reasons why the results from the current experiment differ from those of Experiments 5 and 6. It is possible that the difference reflects a difference in size between the targets and distractors. Specifically, in Experiment 7 the target grandfather clocks were matched approximately, in size, to the distractor bodies. Conversely, in Experiments 5 and 6, the headless body distractors were larger in size than the ladybird targets. Therefore, it appears that when the body distractors are larger in size than the targets (as in Experiments 5 and 6), they capture attention not because of some inherent attentional advantage, but because of their size, i.e., larger surface area, in the scene. Conversely, when targets and distractors are matched for size (as in the Experiment 7), the size of the bodies cannot capture attention. These findings are in line with recent literature observing attention capture by larger objects in a scene (Proulx & Egeth, 2008; Proulx, 2010). However, until this notion is explored further and replicated, this interpretation should be viewed with caution.

Furthermore, given the lack of a body distractor effect here, it is also remains possible that the body distractor effect in Experiments 5 and 6 reflects attention capture by oddness of the headless bodies. This will be examined in Experiment 9.

Another finding worthy of discussion was the presence of a lamp distractor effect. Considering that the lamps were smaller in size (physical area) than the grandfather clocks, this effect cannot be interpreted as the lamps capturing attention because they are larger than the grandfather clocks. Furthermore, the lack of a main effect of target (present or absent) suggests that the lamp distractor effect cannot also be explained in terms of a set-size increase, as regardless of whether the target was present or absent the presence of a lamp slowed responses to the target. Presumably if the lamp distractor effect were the result of a set-size increase, there would have been a main effect of target showing that this effect was only present in target present scenes. An alternative explanation is that because both the clocks and lamps are from the same object category of furniture, the lamp distractor effect may occur because lamps and grandfather clocks share a number of low-level features (e.g., straight edges, similar edged angles etc.). Indeed, it is well established that visual search becomes harder and less efficient when the similarity between target and distractor is increased (Duncan & Humphreys, 1989; Pashler, 1987; Treisman & Souther, 1985).

According to the Guided Search Model of visual attention (Cave & Wolfe, 1989; Wolfe, 1994), each stimulus in a scene is given an activation value depending on the number of features shared with the target. Low similarity therefore means less activation in the visual field and biases the serial stage of search towards the inspection of the target before the other items in the scene. Conversely, high similarity means more features are shared with the target reducing the within-display difference in activation levels. Therefore, regardless of whether the scene is target present or absent, the more salient items there will be in the scene. This will result in the search of non-target items and consequently cause considerable display-size effects. Here search for the target was less efficient in the presence of the distractor lamp, suggesting that the lamps shared more low-level features, e.g., straight edges, with the grandfather clock targets. Hence attention may have been guided to the location of the lamps, because of this target search in the presence of the lamps was harder and led to increased search times in both target present and absent scenes.

In summary, the results from Experiment 7 appear to suggest the presence of a lamp distractor effect, but the absence of a body distractor effect. As such, Experiment 7 provides no evidence in favour of attention capture by human bodies. Instead, the results suggest that body distractor effect demonstrated in Experiments 5 and 6 may have been a consequence of the headless bodies being odd or because of their large size in the scenes. However, given that these conclusions are inconsistent with previous findings of attention capture by silhouette bodies (Downing et al., 2004; Ro et al., 2007) they should be viewed with caution until they can be reliably replicated. Experiment 8 aims to explore this further.

4.2 Experiment 8

4.2.1 Introduction

Experiment 7 offered evidence suggesting that the body distractor effect observed in Experiments 5 and 6 is a consequence of the large size of the bodies, rather than by some inherent attentional processing advantage, capturing attention. Therefore, Experiment 8 aimed to probe the findings of Experiment 7 by using an alternative target, to further explore the relationship between target and distractor size. To achieve this, chair silhouettes were now used as the targets. Chairs are highly familiar objects that are generally symmetrical and viewed in an upright position. Viewers' will have also had substantial visual experience with them. Importantly, chairs are also substantially smaller in size and area (not solid blocks of colour) than the distractor human bodies. Therefore, it was predicted that if the body distractor effect observed in Experiments 5 and 6 reflects attention capture by the large size of the body distractors, then here where the body distractors are larger in size than the target, the presence of the body should slow search for the target relative to when it is absent.

4.2.2 Method

4.2.2.1 Participants

A convenience sample of 40 participants (28F, 12M) aged between 18 – 61 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 40 minutes.

4.2.2.2 Design, Stimuli, Apparatus, Materials and Procedure

The Design, Stimuli, Apparatus, Materials and Procedure were identical to Experiment 7 except here silhouettes of chairs were used as the targets instead of silhouettes of grandfather clocks. These were created from 12 different photographic images of chairs taken from an open source on the internet. In accordance with the other stimuli they were cropped out of their natural outline and scaled to be a realistic size in a scene. A black filter was applied to the images to transform them into silhouettes. Example stimuli illustrating the experimental conditions can be seen in Figure 23.

Target-absent/distractor-absent

Target-absent/body-distractor present

Target-absent/lamp-distractor present

Target-present/distractor-absent

Target-present/body-distractor present

Target-present/lamp-distractor present

Figure 23. An illustration of the scene stimuli for each of the experimental conditions in Experiment 7.

4.2.3 Results

As before, incorrect button presses and none responses were counted as errors and were removed before analysis of the RT data. Median RTs for each of the conditions were calculated for each participant. The means of these medians are presented in Figure 24, along with the mean percentage of errors made in each condition.

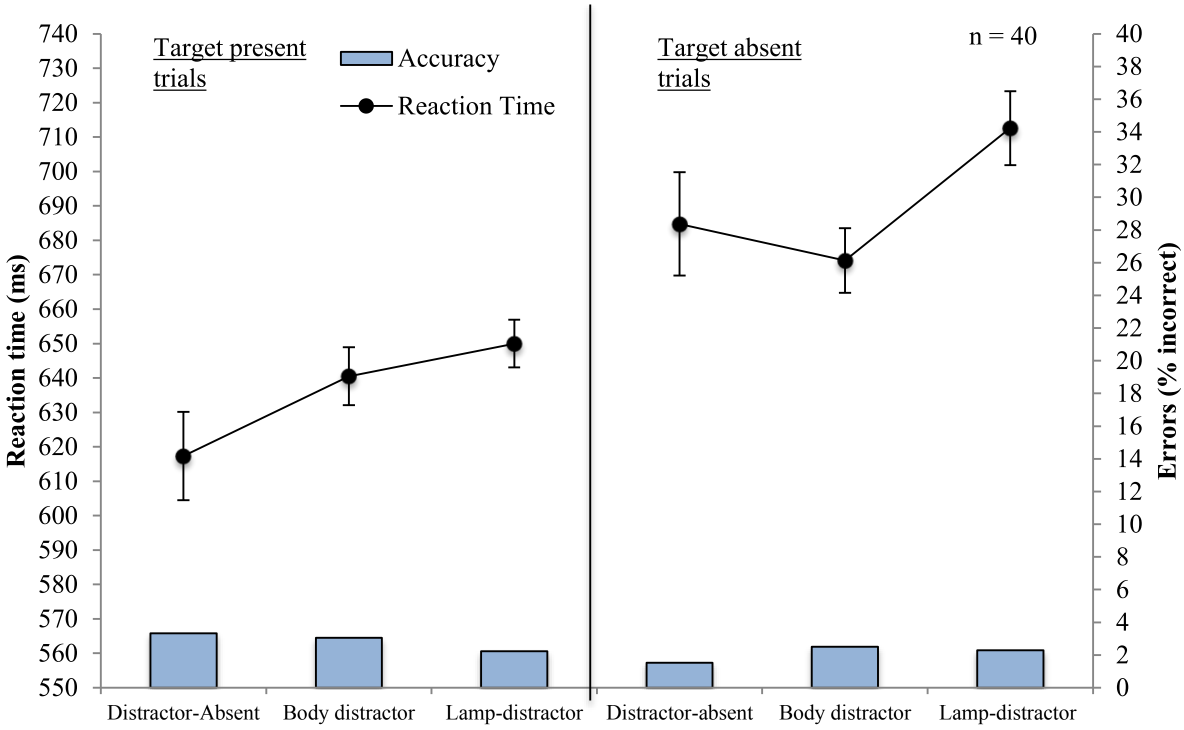


Figure 24. Means of the median correct detection rates and error rates for each condition of Experiment 8 for the target present and absent conditions. Vertical bars presented the within subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 (target: present, absent) × 3 (distractor: absent, body-distractor present, lamp-distractor present) within-subjects ANOVA was conducted on the median RT data. This analysis yielded a significant main effect of target, $F(1, 39) = 39.06$, , $MSE = 178733.13$, $p < 0.01$, $\eta^2 = 0.26$. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2\ (2) = 8.70$, $p < 0.05$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .83$). The results also showed that there was a main effect of distractor, $F(2, 64.75) = 9.88$, $MSE = 24609.36$, $p < 0.01$, $\eta^2 = 0.06$. Again, Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2\ (2) = 8.54$, $p < 0.05$, therefore degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .83$). Finally there was also a significant target × distractor interaction, $F(2, 64.94) = 4.67$, $MSE = 6710.42$, $p < 0.05$, $\eta^2 = 0.02$.

Pairwise comparisons using t-tests with the Benjamini-Hochberg α-error adjustment revealed that for target present scenes, RTs to target-present/distractor-absent scenes ($M = 617.28$) were significantly faster than RTs to target-present/lamp-distractor present scenes ($M = 649.98$; $p < 0.05$, $d = 0.62$) and RTs to target-present/body-distractor present (M = 640.51; $p < 0.05$, d =90.48). There were no significant differences in RTs between the target-present/lamp-distractor scenes (M = 649.98) and target-present/body-distractor present scenes (M = 640.51; $p > 0.05$, d = 0.25).

For target absent scenes, RTs to target-absent/lamp-distractor present scenes (M = 712.63) were significantly slower than RTs to target-absent/distractor-absent scenes (M = 684.75; $p < 0.05$, d = 0.36). However, there was no significant difference between RTs for target-absent/body-distractor present scenes (M = 674.13) and target-absent/distractor-absent scenes (M = 684.75; $p > 0.05$, d = 0.14). RTs for target-absent/body-distractor scenes (M = 674.13) were significantly faster than RTs for target-absent/lamp-distractor scenes (M = 712.63, $p < 0.05$, d = 0.76). All other comparisons were not relevant to the hypotheses and are therefore reported in Appendix 7.

An analogous analysis of the error rates revealed only a main effect of target, $F(1, 39) = 4.70$, $MSE = 35.01$, $p < 0.05$, $\eta^2 = 0.02$, with significantly more errors for target-present trials (M = 2.9%) than target absent trials (M = 2.1%). There was no significant main effect of distractor, $F(2, 39) = 0.82$, $MSE = 0.82$, $p > 0.05$, $\eta^2 = 0.01$. The was no significant interaction target × distractor interaction, $F(2, 78) = 2.37$, $MSE = 18.23$, $p > 0.05$, $\eta^2 = 0.02$.

4.2.4 Discussion

To explore whether the body distractor effect is facilitated by the large size of the a distractor bodies relative to the targets, Experiment 8 used targets (chair silhouettes) that were smaller in size than the body distractors. The results from Experiment 8 demonstrated that in the target present trials the presence of either a distractor body or lamp significantly slowed the visual search for the target relative to when they were absent. In addition, it was also shown that there was no difference in the body distractor effect compared to the lamp distractor effect. The current findings are therefore inconsistent with the results from Experiment 7, which demonstrated the presence of a lamp distractor effect, but the absence of the body distractor effect. Nevertheless, in Experiment 7 the grandfather clock targets and body distractors were all matched approximately in size. However, in Experiments 5, 6 and 7 the body distractors were all larger in size than the chair targets. Taken together, the findings from Experiments 5, 6 and 7 therefore show that the body distractor effect is most likely the result of attention capture by the large size of the bodies relative to the smaller targets. However, given that the headless bodies (Experiment 5 and 6 body distractors) and full-body silhouettes (Experiments 7 and 8 body distractors) have not been directly compared within the same experiment, this conclusion is given very tentatively. In addition, in the target absent trials, the presence of the human body distractor did not slow the search for the target, suggesting that the bodies did not retain attention.

Across both Experiments 7 and 8, regardless of the presence of the target, the presence of the lamp has significantly slowed visual search. This lamp distractor effect is therefore best explained in terms of a similarity in low-level features between the target and the lamps. Specifically, similar to Experiment 7, both the lamps and chairs belong to the category of furniture. Given this, it is plausible that they also share similar features, e.g., straight, angular lines. As previously discussed (section 4.1.4), distractors that are similar to a target can lead to longer latencies in both target present and absent trials (Pashler, 1987).

In summary, the results from Experiments 7 and 8 indicate that the body distractor effect demonstrated in Experiments 5, 6 and 7 are the result of the large size of the bodies relative to the smaller targets, rather than by an inherent attentional advantage of human bodies. The results also suggest that the lamp distractor effect is most likely caused by a similarity in features between the target and distractor.

Chapter 5

5 Target and Distractor Size Contrast, Headless Body Oddity and Attention Capture By Faces

5.1 Experiment 9

5.1.1 Introduction

It would appear that bodies capture attention when they are larger than targets. This has not previously been demonstrated in the literature. Taking that into consideration and given that Experiment 7 was the only experiment to demonstrate no body distractor effect when target and body distractor were matched for size, it was decided to pursue this finding further using the grandfather clocks (Experiments 1 – 3) as targets, and the headless bodies (Experiments 1 – 6) and human full-body silhouettes (Experiments 7 – 8) as distractors. Therefore, all targets and distractors were approximately equal in size. It was predicted that if human bodies capture attention, then the presence of either a headless body or full-body silhouette distractor should slow the search for the target, even when size is controlled. However, if the large size of the body distractors relative to the small targets captured attention in Experiments 5, 6, and 8, then the presence of either a headless or full-body (silhouette) distractor will not slow the search for the target.

In addition, it is also possible that the oddness of the headless bodies may have facilitated the body distractor effect in Experiments 5 and 6. If correct, then the presence of the headless body distractors should slow search for the targets relative to when it is absent, but at the same time, the presence of the full-body silhouettes should not slow search for the target.

5.1.2 Method

5.1.2.1 Participants

A convenience sample of 40 participants (34F, 6M) between 19 – 25 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 40 minutes.

5.1.2.2 Design

A 2 × 3 within-subjects design was employed with two factors: target (present vs. absent) and distractor (absent, headless body-distractor present, full-body silhouette present). The dependent variables (DV) were the reaction time taken from stimulus onset and participant response and participants accuracy. The experiment used an ASP paradigm and the data were analysed using a 2 × 3 analysis of variance. Post hoc analysis (where appropriate) used paired *t* tests with a false-discovery rate correction for multiple comparisons (Benjamini and Hochberg, 1995).

5.1.2.3 Stimuli

The 24 backgrounds (12 indoor, 12 outdoor) and 12 target grandfather clock stimuli were identical to those from Experiments 1 – 3. The 12 body-only stimuli were identical to those used in Experiments 5 and 6. The full-body silhouette stimuli were identical to those from Experiment 7 and 8. For the target present scenes, the target (grandfather clock) was presented in isolation with no distractor present (target-present/distractor-absent condition), or with a headless body distractor in the scene (target-present/body-only-present), or the target was presented with a full-body silhouette distractor in the scene (target-present/full-body present condition). For the target absent scenes, either the scene contained no target or distractor (target-absent/distractor-absent), or the headless body distractor was presented in isolation in the scene (target-absent/body-only present condition), or either the full-body silhouette distractor was presented in isolation in the scene (target-absent/full-body present condition). Example stimuli illustrating the experimental conditions can be seen in Figure 25.

Target-absent/distractor-absent

Target-absent/body-only present

Target-absent/full-body present

Target-present/distractor-absent

Target-present/body-only present

Target-present/full-body present

Figure 25. An illustration of the stimuli for each of the experimental conditions in Experiment 9.

5.1.2.4 Apparatus and Materials

The apparatus and materials for Experiment 9 were identical to that used in Experiments 1 – 8.

5.1.2.5 Procedure

The procedure for Experiment 9 was identical to the one used in Experiments 7 – 8. The participants received 216 trials comprising of 108 target present scenes (36 trials for target-present/distractor-absent scenes, 36 trials for target-present/full-body silhouette distractor present, 36 trials for target-present/body-only distractor present), and 108 target-absent scenes (36 trials for target-absent/distractor-absent scenes, 36 trials for target-absent/full-body silhouette distractor present, 36 trials for target-absent/body-only distractor present). The trials were randomized for each participant.

5.1.3 Results

As in previous experiments, incorrect button presses and none responses were counted as errors and were removed before analysis and included in the accuracy analysis. Median RTs for each of the conditions were calculated for each participant. The means of these medians are presented in Figure 26, along with the mean percentage of errors made in each condition.

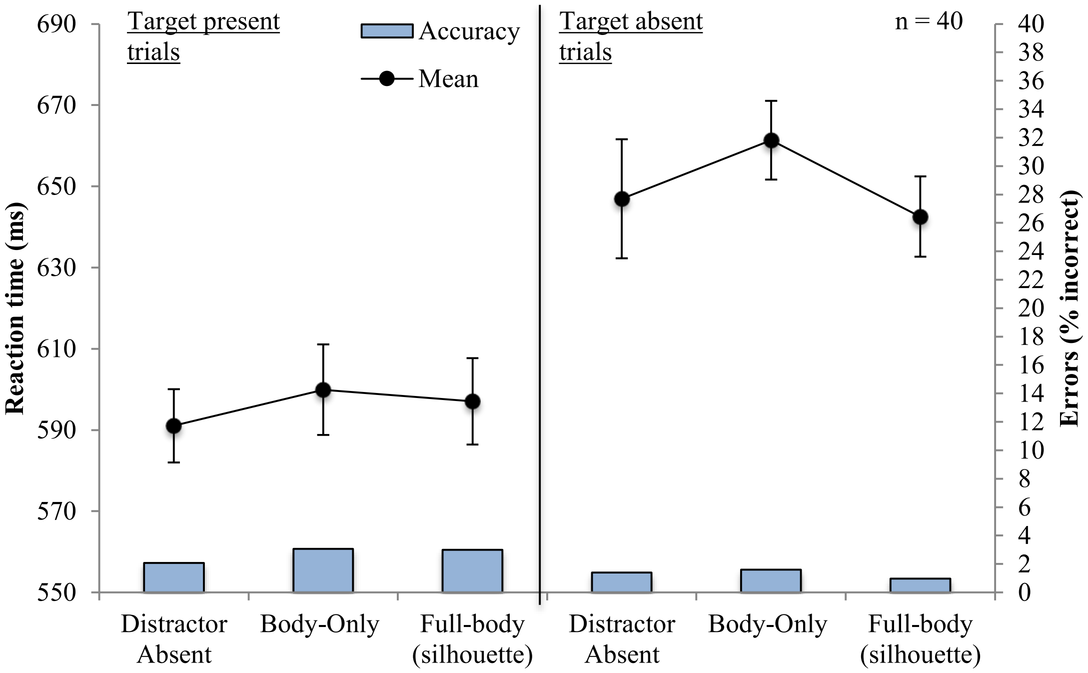


Figure 26. Means of the median correct detection rates and error rates for each condition of Experiment 9 for the target present and absent conditions. Vertical bars presented the within subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 (target: present, absent) × 3 (distractor: absent, body-only distractor, full-body silhouette distractor) within-subjects ANOVA was conducted on the median RT data. This analysis yielded a significant main effect of target, $F(1, 39) = 23.56$, $MSE = 176773.68$, $p < 0.01$, $\eta^2 = 0.26$, with significantly faster RTs for scenes that contained a target ($M = 596.00$) compared to RTs for scenes that did not contain a target ($M = 650.28$). The analysis also revealed a main effect of distractor, $F(2, 78) = 3.74$, $MSE = 6750.59$, $p < 0.05$, $\eta^2 = 0.01$. Pairwise corrected comparisons showed that RTs were significantly faster in scenes that contained a full-body silhouette ($M = 619.81$) compared to RTs for scenes that contained a body-only target ($M = 630.63$; $p < 0.05$, $d = 0.43$). There was no significant difference between distractor absent trials ($M =$

618.99) and full-body silhouette (*M* = 619.81, $p > 0.05$, $d = 0.03$). There was also no significant difference between distractor absent (*M* = 618.99) and scenes that contained body-only targets (*M* = 630.63, $p > 0.05$, 0.32). No significant target × distractor interaction, $F(2, 78) = 0.83$, $MSE = 1305.45$, $p > 0.05$, $\eta^2 = 0.00$.

An analogous analysis was conducted on the error data, this revealed a main effect of target, $F(1, 39) = 10.70$, $MSE = 174.21$, $p < 0.01$, $\eta^2 = 0.08$, with significantly more errors for target present trials (*M* = 2.71%) as compared to target absent trials (*M* = 1.32%). There was no significant main effect of distractor, $F(3, 117) = 1.12$, $MSE = 7.42$, $p > 0.05$, $\eta^2 = 0.01$. There was no significant target × distractor interaction, $F(3, 117) = 1.42$, $MSE = 9.16$, $p > 0.05$, $\eta^2 = 0.01$.

5.1.4 Discussion

Experiment 9 aimed to explore further the relationship between target and distractor size. The data collected from Experiment 9 demonstrates that the time taken to detect the presence of a grandfather clock target was not increased by the presence of either a human full-body silhouette or body-only distractor. This finding is consistent with the results from Experiment 7 and therefore adds further support the hypothesis that when the targets and body distractors are matched for size, the body distractor effect is lost. In turn, this finding adds to cumulative evidence that the body distractor effects in Experiments 5, 6 and 8, where the body distractors were larger in size than the targets, reflects attention capture by the size of the bodies, rather than by some inherent attentional advantage. Taken together, the results from Experiments 5 – 9 therefore provide little evidence in favour of attention capture by bodies. Instead, these findings are in line with previous suggestions that large objects in a scene capture attention irrespective of top-down goals (Proulx & Egeth, 2008; Proulx, 2008).

In addition, it was also predicted that if the headless bodies being odd caused the distractor effect in Experiments 5 and 6 then here their presence should slow search for the target, but that the full-body silhouettes would not. Although the results do not reflect this possibility, a somewhat peculiar finding in this experiment was that regardless of whether the target was present or absent search in scenes that contained a headless body was significantly slower than search in scenes that contained a full-body silhouette. Therefore, although the presence of the headless bodies was not slowing search for the target relative to when they were absent, they were slowing search relative to when the full-body silhouettes were present. This finding could therefore indicate that there are some biases towards the headless bodies, which could in turn reflect the bodies being odd in the scenes (Chen & Reingold, 2001; Wolfe, 2001). In turn, this could suggest that the body distractor effect in Experiments 5 and 6 was partially facilitated the body-only targets being odd in the scenes. However, this interpretation cannot entirely explain the body distractor effect observed in Experiments 5 and 6, as if this were correct, here visual search in presence of the headless body distractors would have been significantly slower than search in their absence. The results here therefore suggest that the body distractor effect in Experiments 5 and 6 can be explained in terms of being both larger than the targets and being unfamiliar stimuli. This was explored further in the final experiment (Experiment 10).

5.2 Experiment 10

5.2.1 Introduction

The results from Experiments 5 – 9 provide strong evidence that the body distractor effect is principally the result of the larger size of the bodies relative to the targets when presented in the scenes. However, this raises an important issue. Faces as visual stimuli were not explored further than Experiment 4. The rationale for this was that the results presented up to that point had been consistent with previous findings of attention capture by faces (e.g., Langton et al., 2008; Ro et al., et al, 2007). However, whilst the face distractor effect in Experiment 4 cannot be attributed to the face distractors being larger than the ladybird targets, it is possible that oddness (unusualness) of the face-only targets appearing to float ‘body-less’ in the scenes may have driven the face distractor effect. Specifically, faces on day-to-day basis are not generally viewed without the body and this factor may have caused the faces to be more salient in the scenes. Likewise, target size notwithstanding, some of this bias for the headless bodies may also be attributable to them being unfamiliar in the scenes – although not enough to facilitate the body distractor effect entirely.

In Experiment 10, full-human bodies (including the depiction of a human face) were used as the distractors. Given that faces and bodies are often perceived together this negated the possibility that the face (or body) will be unfamiliar in the scene and facilitate attention capture. Additionally, grandfather clocks were used as the targets.

The experiments present thus far have established that the body distractor effect for full-bodies appears to be the result of the large size of the bodies relative to smaller targets. Therefore, using targets (grandfather clocks) and distractors (full-bodies) that are approximately the same size eliminates the possibility that any observed distractor effect for the full-bodies (with faces) is the result of target/distractor size difference. This manipulation allowed us to examine whether the distractor face effect observed in Experiment 4 was the result of an attentional advantage or because they were odd stimuli. Therefore, any slowing of visual search for the target in the presence of the full-bodies can only be attributed to the face. It was therefore predicted that, if faces capture attention search for the target should be slower in the presence of the full-body (including the depiction of the face) compared to when it is not present in the scene. In turn, this finding would contribute to cumulative evidence that the human face has some attention bias that captures a viewer’s attention. However, if the face distractor effect observed in Experiment 4 was the result of the oddness of the faces, then the presence of the full-body (including the face) will not slow search for the target relative to when it is absent.

An additional aim of the study was to further examine the lamp distractor effect. The results of Experiments 7 and 8 have consistently demonstrated that the silhouettes of lamps significantly slow the visual search for a target. This effect appears to be the result of a similarity in features between the targets (silhouette grandfather clocks and chairs) and lamp distractors. Specifically, grandfather clocks, lamps and chairs all belong to the same object category of furniture, therefore when presented as silhouettes they may contain a number of similar features. It is therefore possible that presenting lamps as silhouettes removes the features that make them maximally distinct from other objects. As discussed previously, search for a target in the presence of distractors with similar features makes search less efficient (Treisman & Gormican 1988; Wolfe et al. 1989; Duncan and Humphreys 1989; Findlay 1997). This also raises another important issue. At no point have the distractor lamps been presented including their finer details (e.g., colour, internal features). Given that the effect of human body distractors has been examined in both full-detail and in silhouettes, in order to be experimentally consistent the effect of the lamp distractors including finer details should also be examined. Therefore, it was predicted that if presenting the lamps as silhouettes in Experiment 7 and 8 removed features that made them distinct and different from the clocks, then here when they are presented with these features they should not slow search for the targets. On the other hand, if presenting lamps as silhouettes did not remove the features that made them similar to the grandfather clock targets then a distractor lamp effect will still be observed.

In addition, headless bodies were used as a non -face and -lamp distractor. This was included to explore the differences in visual search for the target in the presence of the distractor headless bodies, full-bodies (in colour) and lamps (in colour). The results from Experiment 9 indicated that there might be some biases for headless bodies relative to full-bodies and that this may reflect headless bodies being odd in a scene. It was therefore predicted that if headless bodies slow search for a target because they are odd in a scene, then the presence of a headless body distractor will slow search for the target more than the full-bodies in colour and lamps in colour.

5.2.2 Method

5.2.2.1 Participants

A convenience sample of 40 participants (35F, 5M) between 19 – 44 years were taken from Nottingham Trent University. All of the participants had normal or corrected-to-normal vision. Where applicable, participants were awarded course required research credits upon completion, one credit per 10 minutes. The experiment lasted approximately 40 minutes.

5.2.2.2 *Design*

A 2 × 4 within-subjects design was employed with two factors: target (present vs. absent) and distractor (absent, headless body-distractor present, full-body colour present, lamp-distractor present). The dependent variables (DV) were the reaction time taken from stimulus onset and participants accuracy. The experiment used an ASP paradigm and the data were analysed using a 2 × 4 analysis of variance. Post hoc analysis (where appropriate) used paired *t* tests with a false-discovery rate correction for multiple comparisons (Benjamini and Hochberg, 1995).

5.2.2.3 *Stimuli*

The 24 backgrounds, 12 full-body colour, and 12 target grandfather clock stimuli were identical to those from Experiment 9. The 12 lamps stimuli were identical to those used in Experiment 7 and 8, however, they were presented in their original form including all features, e.g., colour. For the target present scenes, the target was presented in isolation with no distractor present (target-present/distractor-absent condition), or with a headless body distractor (target-present/body-only condition), or the target was presented with a full-body (colour) distractor (target-present/full-body condition), or the target was presented with a lamp distractor (target-present/lamp-distractor condition). For the target absent scenes, the scene contained no target or distractor (target-absent/distractor-absent condition), or the headless body distractor was presented in isolation (target-absent/body-only condition), or the full-body (colour) distractor was presented in isolation (target-absent/full-body condition), or the lamp distractor was presented in isolation (target-absent/lamp-distractor present condition). Example stimuli illustrating the experimental conditions can be seen in Figure 27.

Target-absent/distractor absent

Target-absent/body-only present

Target-absent/full-body present

Target-absent/lamp present

Target-present/distractor absent

Target-present/body-only present

Target-present/full-body present

Target-present/lamp present

Figure 27. An illustration of the stimuli for each of the experimental conditions in Experiment 10.

5.2.2.4 Apparatus, Materials and Procedure

The apparatus, materials and procedure for Experiment 10 were identical to that used in Experiment 9. The participants received 288 trials comprising of 144 target present scenes (36 trials for target-present/distractor-absent scenes, 36 trials for target-present/body-only distractor present, 36 trials for target-present/full-body distractor present, 36 trials for target-present/lamp-distractor present), and 144 target-absent scenes (36 trials for target-absent/distractor-absent scenes, 36 trials for target-absent/ body-only distractor present, 36 trials for target-absent/full-body distractor present, 36 trials for target-absent/lamp-distractor present).

5.2.3 Results

As in the previous experiment, incorrect button presses and none responses were counted as errors and were removed before analysis of the RT data. Median RTs for each of the eight conditions were calculated for each participant. The means of these medians are presented in Figure 28, along with the mean percentage of errors made in each condition.

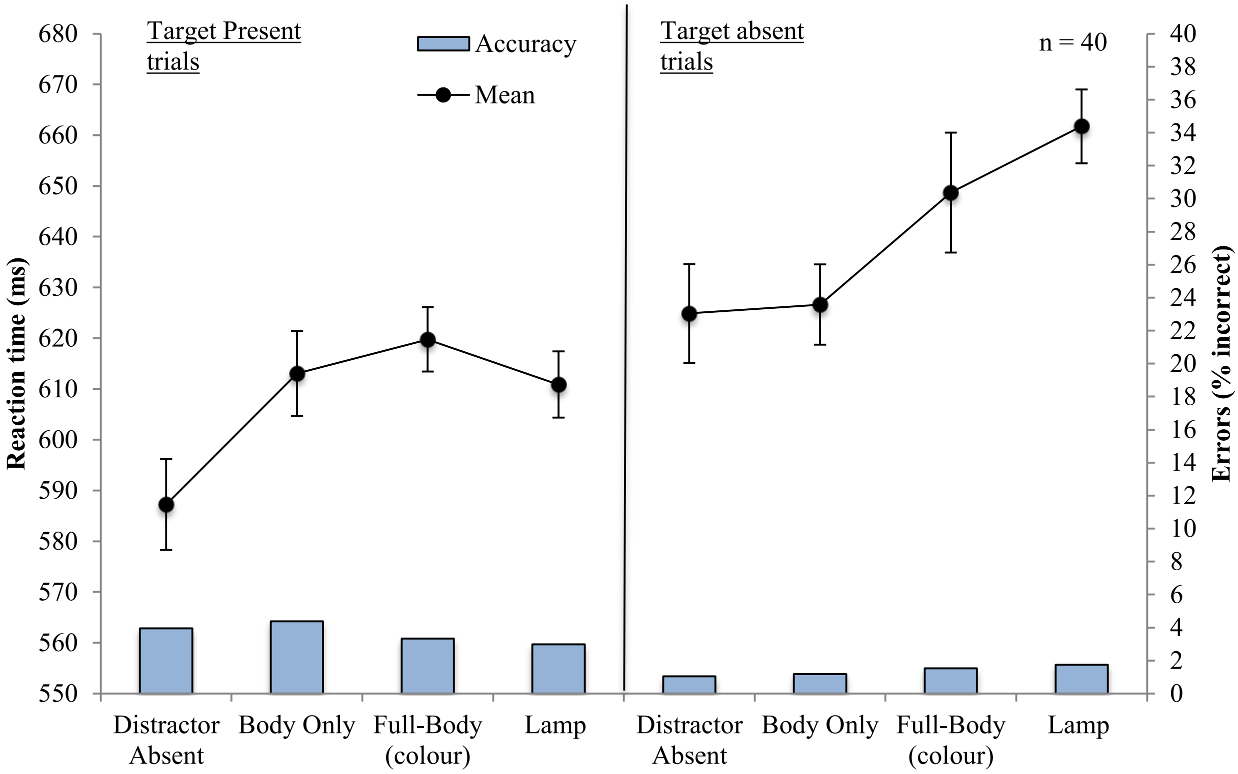


Figure 28. Means of the median correct detection rates and error rates for each condition of Experiment 10 for the target present and absent conditions. Vertical bars presented the within subjects 95% confidence intervals (Cousineau, 2005; Morey, 2008).

A 2 (target: present, absent) × 4 (distractor: absent, body-only present, full-body colour present, lamp-distractor) within-subjects ANOVA was conducted on the median RT data. This yielded a significant main effect of target, $F(1, 39) = 26.03$, $MSE = 85919.66$, $p < 0.01$, $\eta^2 = 0.16$. Mauchly's test indicated that the assumption of sphericity had been violated, $\chi 2$ (5) = 11.60, $p < 0.05$, therefore the degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\varepsilon = .83$). The results showed that there was a significant effect of distractor, $F(3, 117) = 5.17$, $MSE = 19132.74$, $p < 0.05$, $\eta^2 = 0.09$. A significant interaction between target and distractor was also found, $F(3, 117) = 4.40$, $MSE = 4889.86$, $p < 0.01$, $\eta^2 = 0.03$.

Paired comparisons revealed that for target present scenes, RTs for target-present/distractor absent scenes ($M = 587.23$) were significantly faster than RTs for target-present/body-only ($M = 613.04$, $p < 0.05$, $d = 0.06$), target-present/full-body ($M = 619.80$, $p < 0.05$, $d = 0.87$), and target-present/lamp-distractor scenes ($M = 610.89$, p

< 0.05, $d = 0.58$). RTs for target-present/body-only scenes ($M = 613.04$) were not significantly different from RTs for target-present/full-body ($M = 619.75$, $p > 0.05$, $d = 0.18$), and RTs for target-present/lamp-distractor ($M = 610.89$, $p > 0.05$, $d = 0.05$) scenes. There was also no significant difference between RTs for target-present/full-body scenes ($M = 619.75$) and target-present/lamp-distractor scenes ($M = 610.89$, $p < 0.05$, $d = 0.24$).

For the target absent scenes, RTs for the target-absent/distractor-absent scenes ($M = 624.90$) were significantly faster than RTs for the target-absent/lamp-distractor scenes ($M = 661.75$, $p < 0.05$, $d = 0.80$). RTs for target-absent/distractor-absent scenes ($M = 624.90$) were not significantly different from RTs for target-absent/body-only ($M = 626.64$, $p > 0.05$, $d = 0.04$) and target-absent/full-body ($M = 648.70$, $p > 0.05$, $d = 0.40$) scenes. In addition, RTs for target-absent/body-only scenes ($M = 626.64$) were significantly faster RTs for target-absent/lamp-distractor scenes ($M = 661.75$; $p < 0.05$, $d = 0.78$). There were no significant differences observed between the RTs for target-absent/lamp-distractor ($M = 661.75$) and target-absent/full-body scenes ($M = 648.70$; $p > 0.05$, $d = 0.23$). RTs for target-absent/body-only scenes ($M = 626.64$) were significantly slower than RTs or target-absent/full-body scenes ($M = 648.70$; $p < 0.05$, $d = 0.35$). All other comparisons were not relevant to the hypotheses and are therefore reported in Appendix 8.

An analogous analysis was conducted on the error data, this revealed a main effect of target, $F(1, 39) = 27.38$, $MSE = 420.14$, $p < 0.01$, $\eta^2 = 0.15$, with significantly more errors for target present trials ($M = 3.67\%$) as compared to target absent trials ($M = 1.37\%$). There was no main effect of distractor, $F(3, 117) = .36$, $MSE = 2.67$, $p > 0.05$, $\eta^2 = 0.00$. There also significant target × distractor interaction, $F(3, 117) = 2.54$, $MSE = 16.84$, $p > 0.05$, $\eta^2 = 0.02$.

5.2.4 Discussion

Experiment 10 showed that the presence of the body-only, full-body and lamp distractors all increased the time taken to detect the presence of the grandfather clock targets relative to when they were absent. In conjunction with the results from

Experiment 4, the finding that the presence of a full-body (including the face) distractor slows the visual search for the target is in line with other recent findings that attention is captured by human faces (e.g., Langton et al., 2008; Ro et al., 2007). In turn, this finding also rules out an explanation for the face distractor effect in Experiment 4 based on the faces capturing attention because they were odd stimuli in the scenes.

In the target present trials, the finding that the presence of the headless body distractors slowed visual search for the targets was inconsistent with the results of Experiment 9. However, they did not slow search for the target more than the lamps or full-bodies, as was the case in Experiment 9. Considering that previous experiments had shown that the large size of the body distractors relative to the smaller targets captured attention, the finding of a headless body distractor effect for size was particularly unexpected, especially since the headless bodies were approximately the same size as the targets. One potential explanation for this finding it that it may reflect headless bodies being odd in scenes. This was discussed in Experiment 9, where the results indicated that oddness of headless bodies drove some biases towards them.

Analysis of the target absent trials revealed that neither the full-bodies (colour) nor headless bodies slowed search for the target when they were absent. Research has indicated that faces and bodies may retain attention more than other visual stimuli (Bannerman et al., 2010; Ro et al., 2007; Bindemann et al., 2005b). However, the face and headless body distractor effect demonstrated here was unlikely to be the result of an attention retention, as presumably the distractor face and headless body distractor effects would have also been apparent in the target absent scenes.

In contrast, the distractor lamps slowed visual search in both target present and absent scenes. Therefore, in line with the suggestions from Experiments 7 and 8, the most likely explanation for this effect is that regardless of whether presented as silhouettes or with their finer details, the lamps and target grandfather clocks share a number of similar low-level features. As discussed in Chapter 4, this is most likely because they are exemplars from the same object category of furniture. Further, research has shown that search for a target in the presence of distractors with similar features makes search less efficient, hence the slower search for the targets (grandfather clock or chair) in the presence of the lamps (Treisman & Gormican 1988; Wolfe et al. 1989; Duncan and Humphreys 1989; Findlay 1997).

5.3 General Discussion

In summary, Experiments 9 and 10 have produced a number of interesting findings. The key finding was further support for the notion that body distractors (headless bodies and body silhouettes) are matched to the size of the target, they do not slow search for the target. This finding strongly suggests that the large size of the bodies relative to the target captures attention.

An important finding of Experiment 9 was that although the presence of headless bodies did not slow search for the target relative to their absence, they did slow search for the target more than the full-body silhouettes. This can be explained in terms of some potential biases towards the headless body distractors, which is most likely the result of the headless bodies being odd in the scenes. This was supported in Experiment 10 when the presence of the headless bodies slowed search for the target. Experiment 10 also demonstrated that visual search for a target in the presence of a full-body containing the depiction of a face is significantly slower than when the full-body was absent. As targets and distractors were matched for size this effect cannot be attributed to the size of the body, instead, it is in line with evidence suggesting attention capture by faces (Langton et al., 2008; Ro et al., 2007).

The experiments in this chapter have achieved a number of things. First they provide support for the notion that body distractor effect for full-body silhouettes is the result of the bodies capturing attention owing to their large size in the visual scene. Second, the data collected suggest that whilst the headless body distractor effect is primarily owing to their larger size in the scene, this effect may also be driven by the oddity of their appearance. Further, they have demonstrated that human faces capture attention.

Chapter 6

6 General Discussion

6.1 Introduction

The central aim of this thesis was to investigate whether human faces and bodies have an attentional priority processing advantage when presented in natural scenes. The introduction reviewed previous studies that are of importance to this thesis and identified a number of limitations. Although a considerable number of studies have examined and established that there are attentional biases for faces, there is comparatively less behavioural evidence in favour of a similar attentional bias for human bodies. This is surprising given the social significance and neural specificity of the human body as a visual stimulus. Furthermore, most previous investigations of preferential attention towards faces and bodies have presented the stimuli in simple displays, namely uniform colour backgrounds. Consequently, any findings of advantages for faces or bodies are confined to these simple conditions and are therefore limited in what they can say about attention capture in the real world (Bindemann et al., 2010). This thesis presented a novel and original approach to these issues. Over a series of 10 experiments, we measured the detection rates to face and body targets in a visual search task and the effect of task-irrelevant face and body distractors in an additional singleton paradigm (ASP).

6.1.1 Summary Overview of Findings

Recent evidence (Bindemann et al., 2010) demonstrated that in natural scenes human bodies are detected as fast as human faces, suggesting that like faces, bodies might also have a (putatively attentional) detection advantages in natural scenes. Experiment 1 (Chapter 2) aimed to replicate this finding, but also to extend it by comparing response times for human targets (face-only, body-only and full-body) to analogous grandfather clock targets (clock face-only, body-only and full-body) in natural scenes. The results from Experiment 1 were not consistent with Bindemann et al. (2010) instead they demonstrated that the bodies (body-only and full-body targets) were detected faster than faces. As such, it appeared that the bigger in size the target, the easier (faster) it is to detect.

The results of Experiment 1 (Chapter 2) also revealed that there were no detection advantages for human targets relative to the grandfather clock targets. However, Experiment 1 presented targets in a blocked design, which might have led to an attentional set that equalled out target type (human, clock) detection (Bruder & Ribeiro-do-Valle, 2009). In an attempt to address this issue, Experiment 2 (Chapter 2) used a randomized design whereby human and clock targets were matched on the basis of target part in each block (face-only, body-only, full-body only). Consistent with Experiment 1, Experiment 2 revealed that irrespective of type, the full-body targets had a detection advantage over both face-only and body-only targets, whilst body-only targets had an advantage over face-only targets. However, contrary to Experiment 1, human targets were detected faster than the grandfather clocks. This finding is consistent with the idea that human targets are processed faster than other objects in natural scenes.

The discrepancy in findings between Experiments 1 and 2 can be explained in terms of the way in which the targets were presented independently. However, it could also be argued that some other factor (e.g., low-level image-based properties) might have resulted in the detection advantages for human targets observed in Experiment 2. Therefore in Experiment 3 (Chapter 2) we attempted to replicate Experiment 2 but with all of the targets inverted (rotated 180 degrees). Detection rates for inverted clock targets were compared to detection rates for inverted human targets. The results from Experiment 3 demonstrated that inverting the targets eliminated the detection advantages for the human targets, suggesting that low-level image properties cannot explain the advantages observed in Experiment 2. Additionally, the detection rates for the inverted human targets were also compared to detection rates for the upright human targets from Experiment 2. Importantly, this revealed no inversion effect for face-only, body-only or full-body targets, indicating that the human target detection advantages in Experiment 2 are not the result of some inherent processing advantage.

Considering the uncertainty in the results of Experiments 1 – 3 (Chapter 2), the remainder of the thesis used the additional singleton paradigm (ASP), which is considered to be a more stringent test of attention capture (Theeuwes, 1991, 1992, 1994, 1996). In Chapter 3, the thesis explored whether the presence of task-irrelevant body-less faces (Experiment 4) and task-irrelevant face-less bodies (Experiment 5) would slow visual search for a ladybird target in both natural and uniform grey backgrounds. It was predicted that such a finding would support the notion of attention capture by the distractor stimulus (face or body, respectively). Uniform grey backgrounds were included to address issues relating to low-level contrast. It was therefore predicted that if low-level factors were facilitating attention capture, then search in the grey scenes should be faster than search in the natural scenes.

Experiment 4 indicated that in both natural and grey scenes the presence of a task-irrelevant face stimulus significantly slowed the search for the target, relative to when the face was absent (referred to as the face distractor effect). Similarly, in Experiment 5 the presence of a task-irrelevant body significantly slowed the search for the target in both scene types (referred to as the body distractor effect). Taken together, these findings indicate that attention may be captured by the face and body.

In both Experiments 4 and 5 (Chapter 3), analysis of the target absent trials revealed significantly slower search times when a face or body was present compared to when they were absent, suggesting that the faces and bodies may also retain attention. In addition, in Experiments 4 and 5 (Chapter 3), responses in grey scenes that contained a target were significantly faster than natural scenes that contained a target, suggesting that some low-level property of the target was facilitating faster target detection. However despite this, the face and body distractor effect was still observed.

Both Experiments 4 and 5 (Chapter 3) appear to show attention capture and retention by human faces and bodies. However, these findings may have reflected the time costs of serial search examining both a target and a distractor and background, compared to search only having to examine the one target. To address this problem, in Experiment 6 (Chapter 3) all target or distractor present scenes contained two objects. The results showed that the presence of a distractor body continued to slow the visual search for a ladybird target, compared to scenes that contained two ladybird targets. This tentatively suggested that the body distractor effect observed in Experiments 4 and 5 was not the result of the serial search time costs incurred when examining both target and distractor. In addition, there was no difference between the natural and grey scenes that contained a target. This suggests that an increase in visual load in the grey scenes had slowed target processing and that low-level factors of the target were no longer facilitating its detection.

Although the findings from Chapter 3 (Experiment 4 – 6) appeared to be consistent with the notion of attention capture by bodies (i.e., bodies slowed the search for a target) there were a number of experimental shortcomings that were addressed in Chapter 4. First, the body distractors used in Experiments 5 and 6 were headless bodies. However, bodies are rarely seen without an accompanying head, as such they may be unfamiliar by which we mean unusual in a sense that they deviate from their normal appearance and therefore, odd. Previous research has shown that search becomes less efficient in the presence of odd distractors (Shen & Reingold, 2001; Wolfe, 2001; Wang et al., 1994), thus by extension the headless bodies may have slowed search because they were odd stimuli in the scene. To avoid this problem, full-body silhouettes were used as body distractors in Experiment 7 (Chapter 4).

Second, there was no comparative control distractor baseline in Experiments 4 – 6 (Chapter 3). Although the body distractor effect indicates attention capture by bodies, the distractor effect was not compared to the effect of a non–body distractor stimulus. Without this comparison, it was impossible to determine whether the distractor effect is unique to bodies. To explore this, in Experiment 7 (Chapter 4) silhouetted lamps were used as the control distractor. It was predicted that if bodies capture attention, then their presence should slow search for the target more than the presence of the lamp distractors.

Third, the ladybird targets used in Experiments 4 – 6 (Chapter 3) were all smaller in size than the bodies. Previous evidence has shown that large objects capture attention in a scene (Proulx & Egeth, 2008; Proulx, 2010). Therefore, silhouette grandfather clocks, which were matched in approximate size to the full-body silhouettes, were used at the targets in Experiment 7 (Chapter 4) to control for object size differences.

The findings from Experiment 7 (Chapter 4) revealed that the presence of the human body distractors did not slow the search for the grandfather clock targets. This led to the suggestion that the body distractor effect seen in Experiments 5 and 6 (Chapter 3) could have been caused by attention capture by the large size of the distractors relative to the small targets, or because of oddness of the headless bodies. Conversely, the presence of the lamp distractors slowed search for the target compared to when it was absent. Given, that the grandfather clock targets and distractor lamps belong to the same object category of furniture and were both presented in silhouette, the presence of the lamp distractor effect most likely reflects a similarity in features between the targets and distractors (Duncan & Humphreys, 1989; Pashler, 1987; Treisman & Souther, 1985).

Experiment 8 (Chapter 4) was a replication of Experiment 7 (Chapter 4), using a target that was smaller in size than the body distractors. Specifically, the silhouette of a chair was used (the distractors remained the same as in Experiment 7: full-body silhouette and lamp silhouette). The results from Experiment 8 demonstrated that the inclusion of both the lamp and full-body silhouette distractors significantly slowed the visual search for the chair targets. Taken together, it appears that the most likely explanation for the body distractor effect in Experiments 5, 6 (Chapter 3) and 8 (Chapter 4) is that the large physical size of the bodies captures attention in a scene, rather than by some inherent attentional advantage.

In an attempt to validate the above conclusion, in Experiment 9 (Chapter 5) the targets (grandfather clocks) were matched for approximate size with the distractor full-body silhouettes and headless bodies. Although the evidence for the body distractor effect thus far appears to be consistent with a target and distractor size relationship, at no point had the headless bodies been compared directly the full-body silhouettes or matched for size with the target. It was therefore possible that the body distractor effect in Experiment 5 and 6 (Chapter 3) reflected headless bodies being odd in the scenes. The results showed that neither the full-body silhouettes nor headless bodies slowed the search for the targets compared to their absence. However, visual search in the presence of the headless bodies was significantly slower than search in the presence of the full-body silhouettes, suggesting that there may be some biases for headless bodies. Although perhaps not enough to drive the body distractor effect entirely, it most likely reflects the oddness of the bodies.

Experiment 10 (Chapter 5) returned to the notion of attention capture by human faces by using a full-body distractor, which included the depiction of the human face. To be experimentally consistent, headless bodies and lamps in colour were also included as distractors. The results showed that when present and relative to distractor absent trials, all of the distractors slowed target search.

The body distractor effect for the full-bodies (including face) is consistent with previous reports of attention capture by faces. However, the finding of a headless body distractor effect was surprising and may further suggest that there is bias towards headless bodies.

6.1.2 Summary Conclusions

In light of the findings discussed above, the following three main conclusions are drawn. First, human faces capture attention in both natural and grey visual scenes. This appears to be because of an attentional priority processing bias for faces. Second human bodies do not show priority processing over other objects in either grey or natural scenes. Any biases or detection advantages observed for body targets are because they larger in size than other objects or, if headless, because they are odd stimuli in that scene. Third human full-body targets, which are perceived on a day-to-day basis, capture attention partly because they include face and partly because they are large objects in the scene. The remainder of this chapter will consider these main conclusions in more depth and in relation to previous findings, whilst also suggesting future directions for exploration.

6.2 Attention Capture By Faces

The findings of Experiments 2, 3 (Chapter 2), 4 (Chapter 3) and 10 (Chapter 5) are consistent with a substantial amount of empirical evidence showing that faces preferentially capture attention (e.g., Hershler & Hochstein, 2005; Lewis & Edmonds, 2003, 2005; Mack et al., 2002; Shelley-Tremblay & Mack, 1999; Vuilleumier, 2000; Bindemann et al., 2007; Langton et al., 2008; Ro et al., 2007; Ro et al., 2001; Theeuwes & Van der Stigchel, 2006). Experiment 2 showed that human face-only targets were detected faster than comparable objects (grandfather clocks) when presented upright in a visual scene. However, the lack of an inversion effect for the inverted face-only targets from Experiment 3 (Chapter 2) compared to the upright face-only targets from Experiment 2 (Chapter 2) raises some important questions. As previously noted, the inversion effect for faces has been taken to suggest that inverted faces do not attract attention than other objects (Yin, 1969; Reed et al., 2004; Diamond & Carey, 1986; Farah et al., 1995; McKone et al., 2007; Yin, 1969). Therefore, given that there were no differences in the detection between the upright and inverted face targets, it could be argued that detection advantages observed for upright face-only over clock targets in Experiment 2 (Chapter 2) were not because of an attentional priority processing. However, evidence suggests that the face inversion effect may be dependent upon the task and the stimuli used (Garrido et al., 2008).

Indeed, whilst the evidence is strong for the face inversion effect in recognition tasks (e.g., Reed et al., 2003, 2005), the evidence for the face inversion effect in face detection tasks has been mixed. For instance, using a face detection paradigm, Purcell and Stewart (1986) presented participants with inverted and upright masked faces on uniform colour backgrounds. They asked participants to locate the face target, finding that performance for face detection was faster and more accurate when the face was in the upright position relative to the inverted position. Similarly, Garrido, et al. (2008) observed large inversion effects for faces in a detection paradigm when the face stimuli were greyscale photographs, which were degraded with noise.

However, several other studies presenting face stimuli on natural backgrounds have demonstrated smaller inversion effects for face detection than for face recognition (Lewis & Edmonds, 2005; Rousselet et al., 2003; VanRullen, 2006). It has therefore been suggested that the small inversion effects in these studies are a result of the stimuli used (Garrido et al., 2008). For instance, when faces are presented in the natural scenes there are several contextual cues available, i.e. backgrounds, skin colour, and external facial features, which might lead to good performance, hence small or no inversion effects (Lewis & Edmonds, 2003, 2005; Rousselet et al., 2003; VanRullen, 2006). Specifically, when these contextual cues are present, participants may focus more on these than the target to aid detection, resulting in faster detection rates for the inverted faces (Robbins & Coltheart, 2012). However, when these are not present, viewers are forced to rely on the face and its configuration, hence the presence of the face inversion effect (Garrido et al., 2008; Purcell & Stewart, 1986). Therefore, given that in Experiments 2 and 3 (Chapter 2) the face-only targets were presented in natural scenes containing a substantial number of contextual cues (e.g., hair, backgrounds, etc.), the lack of a face inversion effect in Experiment 3 is explainable in terms of the participants using non-face cues to aid detection of the inverted face-only targets. Critically, this interpretation implies that the detection advantages for human face-only targets in Experiment 2 (Chapter 2) can be explained in terms of an attentional priority processing for faces, adding support to the notion that faces preferentially capture attention.

To our knowledge no other research has directly examined the presence that contextual cues has on the face inversion effect. Further investigation might provide additional insight into the conditions of the face inversion effect. One way to achieve this would be to establish whether the presence of contextual cues, such as the backgrounds used here, negate the face inversion effect. To do this, participants could be presented with upright and inverted faces in both natural backgrounds and uniform colour backgrounds. It is hypothesized that if participants focus on contextual cues to aid detection, then the face inversion effect should only be present in the uniform colour backgrounds. The experiment outlined here could provide further insight into the conditions under which upright and inverted faces are detected in natural scenes.

The conclusion that faces preferentially capture attention appears to receive further support from Experiments 4 (Chapter 3) and 10 (Chapter 5), where the inclusion of a face significantly slowed the visual search for a target. Importantly, the data from Experiment 10 suggested that face distractor effect in Experiment 4 was not caused by the body-less faces appearing odd in the scenes. As previously discussed, evidence from visual search has shown that search for a target is slowed by the presence of odd distractors (Shen & Reingold, 2001; Wolfe, 2001; Wang et al., 1994). As such, it is possible that the faces slowed search for the targets in Experiment 4 because they were odd. To control for the possibility, in Experiment 10, face distractors were presented as part of the full-bodies. Given that Experiments 5 – 9 established that the large size of the body distractors relative to the smaller targets drives the body distractor effect, Experiment 10 also used size-matched grandfather clocks as the targets. This manipulation negated the possibility that the full-body distractor effect could be explained in terms of attention capture by the large size of the full-bodies. Instead, the full-body distractor effect can only be attributable to attention capture by the face depicted in the full-body. As such, the presence of a full-body distractor effect in Experiment 10 provides evidence in favour of the face distractor effect observed in

Experiment 4 being the facilitated by preferential attention towards faces. In turn, this suggests that the face distractor effect observed in Experiment 4 (Chapter 3) was not because of the face distractors being odd in the scenes.

Another finding worth of discussion was that in Experiment 10, the presence of the full-body (including the face) did not slow search more than the presence of a lamp distractor slowed search for the target. It had previously been predicted (Chapter 5) that if an object captures attention, then it should slow search more relative to a control distractor. Therefore, it could be argued that the faces depicted within the full-bodies were not capturing attention relative to the lamp distractors. However, as shown in Experiments 7, 8 and 10, the lamp distractors most likely slowed the visual search in both target present and absent scenes because they are from the same object category of furniture and therefore share a number of low-level features with the targets, e.g., straight angular edges. According to the Guided Search theory of attention (Wolfe, 1994), if a distractor shares many features with a target, they will be given a high activation within the visual field. Therefore, in both target present and absent trials they will be an area of high activation and attention may be mistakenly guided towards them. Once there, they may require longer inspection times before they can be rejected and attention can search the rest of the scene. This can therefore explain presence of the lamp distractor effect in both target present and absent scenes in Experiments 7, 8 and 10 (Nickerson, 1965; Scialfa & Thomas, 1994). One way to avoid this would be use distractors that are maximally different from the targets. For example, using the same paradigm, if the target was the grandfather clock, the non –face and –body distractors could be a tree or bush (plant). These do not share many features with a grandfather clock. As such, it would be predicted that there would be no tree or bush distractor effect and that the face would slow search for the target more than the presence of the tree or bush.

Important to note is that the above conclusions suggests that in the context of Experiment 2 (Chapter 2), the detection advantages for the human full-body targets over the clock-targets were primarily the result of the presence of the face. However, in Experiment 2, the human full-body targets were also detected faster than the face-only targets. Therefore, although full-body target detection advantages over the clock targets can be explained by attention capture by faces, advantages for the full-body targets over the face-only targets cannot solely be explained in the same terms. If they were, then there would have been no difference in the detection of the human face-only and full-body targets. In line with Bindemann et al.'s (2010) suggestion, it could therefore be argued that there are some attentional biases towards the human body that contribute to the detection advantages of the full-body targets. However, evidence from Experiments 5 – 10 strongly suggests that detection advantages for human bodies are the result of their large size in a scene. Taking this into consideration and given that at the same viewing distance the full-body targets are larger than the face-only targets, the most parsimonious explanation of the faster detection times of full-human body targets over the face-only targets is that they are larger in size.

It is therefore suggested that full-body targets are detected faster in scenes partly because they contain a face (which has a preferential processing status), and partly because they are large in scenes. Notably, this conclusion conflicts directly with conclusions from previous research (Bindemann et al., 2010). Specifically, Bindemann et al. argued that full-body targets are detected rapidly in a scene because both the face and body contribute equally to their detection. On this basis Bindemann et al. tentatively suggested that the rapid detection of the body targets (body-only and full-body) was indicative of an attentional priority bias for bodies. The results presented here in Experiments 2 (Chapter 2), 4 (Chapter 3) and 10 (Chapter 5) do not support this suggestion. However, unlike in Experiments 1 and 2 (Chapter 2), the detection rates for Bindemann et al.'s human targets were not compared against the detection rates of comparable size stimuli. Without comparing the human targets to other stimuli, it is difficult to draw inference from Bindemann et al.'s study about the nature of the detection advantages for any of the human targets. Consequently, the present data provides arguably the most direct evidence that the processes by which a person is detected in a scene are reliant partly on the presence of a face and partly on the relative size of the body in the scene.

Future work might further explore the effect the presence of the face has on processing biases for full-bodies. One-way to do this would be to control for body size. To achieve this identical full-body targets could be used, but with them facing with their backs to the observer (face-away targets) and facing towards the observer. Research has shown that view of the back of the head does not drive efficient search of the face (Burton et al., 2009), and that the body inversion effect is diminished for full human bodies facing away from the camera (Brandman & Yovel, 2012). Therefore by comparing detection rates to face-away targets to detection rates of full-body targets (with face) it is possible to determine the extent to which the face drives the detection of the full-body. The evidence presented throughout this thesis has suggested that faces drive faster detection of the full-body stimuli. It is therefore hypothesized that if the face contributes primarily to the detection of a full-body in a scene, then full-body targets facing the observer will be detected faster than full-body face-away targets. Ideally, these targets should all be presented in natural scenes, in order to maintain ecological validity.

Of course, both the face detection advantages (Experiment 2) and face distractor effect (Experiments 4 and 10) could be accounted for in terms of low-level image based properties of the face images. However, in accordance with previous research (Langton et al., 2008; see section 3.1. for discussion), Experiment 3 (Chapter 2) showed that a comparison between the inverted face-only targets and inverted clock face-only targets removed any advantages for the human targets. Also, as discussed previously, inverting faces is believed to disrupt holistic processing of these stimuli, but maintains low-level image-based properties that might make them salient in a display. As such, detection advantages that are caused by low-level properties are preserved when the faces are inverted (Langton et al., 2008). Therefore, the lack of a detection advantage for the face-only over the clock targets in Experiment 3 (Chapter 2), taken together with evidence from Experiments 2 (Chapter 2), 4 (Chapter 3) and 10 (Chapter 5), is consistent with the idea that faces capture attention.

6.3 The Body Inversion Effect is mediated by the Presence of the Face

Another finding of interest that should be discussed is the lack of an inversion effect between the inverted full-body targets (Experiment 3) and upright full-body targets (Experiment 2). This finding contrasts with previous reports of an inversion effect for full-human bodies that had led to the suggestion the effect represents the specialized processing of upright bodies (Reed et al., 2003; 2006). However, conflicting evidence has shown that the body inversion effect is associated with activation in the face-selective region of the brain (FFA; Brandman & Yovel, 2010) and that the absence of a head, but not other body parts abolishes the body inversion effect (Brandman &

Yovel, 2012; Minnebusch et al., 2009, Yovel et al., 2010). Therefore, it has since been suggested that the body inversion effect is facilitated by the face and therefore generated by face processing mechanisms (Brandman & Yovel, 2012). Given this suggestion, it is plausible that the body inversion effect is not indicative of specialized processing of upright bodies, but rather that of upright faces. In line with this hypothesis, if the face facilitates the body inversion effect then the lack of a full-body inversion effect (Experiment 2 and 3, Chapter 2) is therefore explicable in the same terms as the lack of face inversion effect. Specifically, as discussed above, participants' use contextual cues, such as backgrounds, to aid detection of inverted full-body targets in Experiment 2 (Chapter 2), which resulted in equal detection times for upright and inverted full-body targets. This may suggest that the detection advantages for the full-body targets over clock targets in Experiment 2 (Chapter 2) were attentional. Moreover, taking the evidence that the face mediates the full-body inversion effect, with findings from Experiment 10 that faces attract attention to full-body, we propose that this interpretation provides additional support for our suggestion that the processes by which a person is detected in a scenes is dependent on the presence of the face (and size if the body is large in the scene).

The lack of a headless body inversion effect for the upright (Experiment 2) and the inverted (Experiment 3) body-only targets is also consistent with the idea that the body inversion effect is mediated by specialised face processing mechanisms rather than specialised body processing mechanisms (Brandman & Yovel, 2012). However, in line with the lack of a face and full-body inversion effect it could be argued that the lack of this effect is the result of participants using contextual cues to aid detection of the body-only targets. This interpretation would suggest that headless bodies are subject to an attentional priority processing. However, based on our finding that oddness drives biases towards headless bodies (as further discussed below) this explanation seems unlikely. Supporting this, evidence has also shown that activation in body-selective areas shows poor discrimination of inverted faces and bodies versus upright faces and bodies (indicative of the inversion effect). However, this disappears for headless bodies, again making the above account unlikely (Brandman & Yovel, 2012; Brandman & Yovel, 2010; Minnebusch et al., 2009, Yovel et al., 2010). Altogether, the lack of an inversion effect for full-body and body-only targets, together with previous findings that the face mediates the face inversion effect, provides evidence that the body inversion effect is dependent on the presence of face. This in turn suggests that the body inversion effect is representative of the specialized holistic processing of the face rather than the body.

6.4 Faces retain Attention

The findings of Experiment 4 and 10 provided contrasting results with regard to the relationship between faces and their ability to retain attention. Experiment 4 (Chapter 3) revealed that human faces retain attention when they were body-less. Conversely, in Experiment 10 (Chapter 5) when presented as part of the full-body no such effect was observed. Interestingly, the evidence for attention retention by faces is also mixed in the literature. Langton et al. (2008), for example, found no face distractor effect in the target absent trials, indicating faces did not retain attention. However, Ro et al. (2007) and Bindemann et al. (2005) observed that faces retained attention for longer than other objects. This discrepancy can be explained in terms of methodological differences. Both Bindemann et al (2005) and Ro et al. (2007) used paradigm where attention was first directed towards the face and then was directed elsewhere. For instance, Bindemann et al. (2005) used a go/no-go task where a face was displayed behind the go/no-go stimulus. On go trials, participants were expected to re-orientated attention towards a peripheral cue away from the face. They argued that if faces retain attention responses to the peripheral target should be slowed more by the presence of a face behind the go-stimulus than by other visual stimuli behind the go-stimulus.

Similarly, Ro et al. (2007) used colour-frame distractor singletons to capture attention, whilst participants searched an array for a target frame. Items inside the distractor frames varied, but in order for the target frame to be located, attention had to be disengaged from the distractor frame and re-orientated towards the target frame. When a face was present in the distractor frame, it caused a significant delay on target detection relative to when other items were present in the distractor frames. As such, in both Bindemann et al. (2005) and Ro et al. (2007), attention was orientated first to the face and then disengaged. Conversely, in Langton et al. (2008) faces were task irrelevant and attention was not directed towards them as part of the task. This suggests that faces retain attention only when attention has been directed to them first.

Consistent with this previous research in Experiment 10, where attention was not first directed to the full-bodies, the full-bodies including the face did not retain attention (Bindemann et al., 2005; Ro et al., 2007). Conversely, in Experiment 4 (Chapter 3) faces appeared to retain attention despite the fact that attention was not first directed by the task to the face, thus conflicting with previous research (Langton et al., 2008). Given these findings and the evidence in this thesis that there is a preferential attentional bias for faces, it is possible that attention does not need to be directed towards the face for them to retain attention. Thus, this interpretation is inconsistent with Langton et al. (2008) who showed no attention retention by faces. However, Experiment 4's methodology differs from Langton et al. in several ways. First, Langton et al. presented all their stimuli in rectangle boxes as part of an array. Further, Langton et al.' arrays were presented on uniform colour backgrounds. However, in Experiment 4, face stimuli were presented in varying locations and in natural scenes. Therefore this difference in methodology may reflect the difference in results, whereby some artefact of Langton et al.'s design did not facilitate attention retention by faces.

Moreover, if faces do retain attention, the finding that the full-body including the face does not retain attention is more difficult to explain. One possible explanation is that when faces are attached to the human body as the full-body, some artefact of the human body masks attention retention by faces. This could possibly reflect its large size. Specifically, when presented in a scene attention may be captured by the large size of the body, which consequently masks any difficulty in disengaging attention away from the face (presented in the body). Experiment 8 (Chapter 4) offers some support for this notion, where it was observed that the full-body silhouettes did not slow search in the target absent trials, indicating that they did not retain attention. Although the face was not visible in the full-body silhouettes, the findings do suggest that the full-body silhouettes do not retain attention, even though attention is captured by their large size. Of course, being largely post hoc contemplation, this reasoning is purely speculative. Future research should assess attention retention by full-human bodies, in particular whether the combination of the face and body leads to an absence of attention retention by faces.

6.5 Attention Capture by Human Body Size

The data from Experiments 5 and 6 (Chapter 3) both appeared to provide evidence of attention capture by headless human bodies. These results are therefore consistent with Downing et al. (2004), who observed that briefly presented task-irrelevant bodies were detected, even though participants were engaged in a sustained attention task. It is important to note is that although Downing et al. used a different paradigm (inattentional blindness), to the one used here (additional singleton), in this particular case irrespective of the paradigm, the outcomes are the same, the presence of task-irrelevant body had an effect on the search for a target, which indicated attention capture by the bodies.

On the other hand, the conclusions from Experiments 5 and 6 (Chapter 3) received little support from the findings of Experiments 7 (Chapter 4) and Experiment 9 (Chapter 5). In both Experiments 7 and 9 it was observed that when the target was matched to the size of a body distractor, the presence of a body distractor no longer slowed the visual search for the target. However, similar to Experiments 5 and 6 (Chapter 3), in Experiment 8 (Chapter 4), where the body distractors were larger than the targets, it was observed that the presence of a full-body silhouette distractor significantly slowed the search for the target.

Collectively, it appears that the body distractor effect is only observed when the body distractors are physically bigger in size than the target (Experiments 5, 6 and 8) and not when they are matched approximately for size (see Experiments 7 and 9). It is proposed that this discrepancy reflects attention capture by the large size of the body distractors relative to the targets. This interpretation also gains support from the unexpected finding that throughout Experiments 1 – 3 (Chapter 2), target detectability was related to target size. That is, the bigger body targets (body-only and full-body) were detected faster than the smaller face-only targets, and that the full-body targets were detection faster than the body-only targets. To reiterate the point, based on recent previous findings (Bindemann et al., 2010), it had been predicted that if bodies have attentional processing advantages then body-only targets should be detected as fast as face-only targets in Experiments 1 – 2. The findings from Experiments 1 – 3 (Chapter 2) did not however reflect this prediction. Bindemann et al. (2010) had acknowledged the possibility that target size was a characteristic that potentially enabled body-only targets to be detected as fast as face-only targets. However, as previously discussed, they dismissed this on the basis that their body-only targets were detected as fast as face-only targets, suggesting that detection rates did not get faster with increasing target size.

The methods used in Experiments 1 – 3 (Chapter 2) were not very different from the ones used by Bindemann et al. (2010). However, Bindemann et al. (2010) presented their stimuli in natural scenes where their body stimuli (body-only and full-body) were often partially occluded by furniture and other objects. Conversely in Experiments 1 – 3 (Chapter 2) the body stimuli were always fully visible in each scene. As a result, in Bindemann et al. when the body-only targets were partially occluded they were, in effect, smaller sized targets. Given this, it is possible that the partial occlusion of body stimuli in Bindemann et al. (2010) masked any potential relationship between target size and detectability, essentially equalling out the detection rates for the face-only and body-only targets. Note, however, that although the full-bodies were also partially occluded in Bindemann et al.'s scenes, they were still detected faster than the face-only and body-only target. These targets, however, also contained a face that was never occluded in any scene. Therefore, even if the full-body were partially occluded, they would have still been larger in size than the body-only targets hence detected faster. Importantly, the current research has an advantage over Bindemann et al.'s (2010) research. Specifically, in Experiments 1 – 3 (Chapter 2) the backgrounds, the target stimuli and stimuli locations were all completely randomised. This prevented any biases towards any of these stimuli and their spatial locations. However, Bindemann et al. (2010) pre-experiment superimposed their human bodies in the scenes therefore having a limited stimuli set, which could have lead to biases towards certain targets.

The above interpretation strongly suggests that in Experiments 1 – 3 (Chapter 2), the advantages for the body-only over the face-only targets was a consequence of their larger size. In turn, this supports the notion that the body distractor effect demonstrated in Experiments 5, 6 (Chapter 3) and 8 (Chapter 4) was also the result of attention capture by the large size of the body distractors relative to the smaller targets. Although this finding has not been observed for human bodies as visual stimuli before, this conclusion is consistent with evidence suggesting that attention is captured by large objects in a scene. Specifically, using an additional singleton paradigm, Proulx and Egeth (2006;

Experiment 2) demonstrated that the search for a vertical bar target in an array of horizontally orientated bars was significantly slowed by the presence of a horizontally orientated bar that was larger than the other bars. Thus, it appears that regardless of the top-down goals of an observer (i.e., the task demands), the presence of a large object (including human bodies) in a scene can automatically capture attention. There is also comparable research demonstrating that there are detection advantages for larger faces over smaller faces when presented in natural scenes (Burton et al., 2009). As such, examining the extent to which bodies capture attention because of their size would be an interesting focus of further investigation.

6.6 Detection Advantages for oddness

Whilst it appears that attention is captured by the large size of the bodies in the scenes, Experiment 10 (Chapter 5) produced an unexpected finding. Specifically, the presence of the headless body distractors, which were the same size as the grandfather clock targets, slowed the visual search for the targets. This finding could be taken as evidence in favour of attention capture by human bodies. However, given that the same effect was not observed in Experiment 9 (Chapter 5), where the headless body distractors were also size-matched to the grandfather clock targets, it seems likely that some other factor facilitated the effect. A possible explanation of this finding is that there are some biases towards headless human bodies. This interpretation would therefore explain why in Experiment 9 (Chapter 5) the headless body distractors slowed visual search relative to search in the presence of the full-body silhouette distractors.

Given that they are not generally perceived without the head (face), one explanation for biases towards headless bodies is that they may represent odd stimuli. This interpretation is congruent with literature demonstrating that odd objects in a scene will attract attention (Shaw & Skolnick, 1994; Mitchell, Livosky & Mather, 1998; Shen & Reingold, 2001; Wolfe, 2001; Wang et al., 1994; Becker, Lubin & Pashler, 2007). For example, Becker et al. (2007) showed that objects that deviate from their normal appearance attract faster and more first eye fixations than normal objects. Interestingly, whilst controlling for low-level image saliency, they also demonstrated that a person’s hand that had been changed from flesh to green was fixated on faster and for longer than when the hand was flesh coloured. Thus, oddness of the headless bodies may produce some biases towards them as distractors. Crucially, if this were a robust effect, as well as slowing search for the full-body silhouette distractors, the presence of the headless bodies would have slowed the search for the grandfather clock targets in Experiment 9 (Chapter 5).

The above interpretation also has implications for the findings from Experiments 5, 6 (Chapter 3) and Experiment 9 (Chapter 5). Specifically, that the headless body distractor effect observed in these Experiments can partly be explained in terms of their large size relative to the targets and partly explained by attention capture because they are odd. In turn, this provides an appealing explanation for the detection advantages observed for human body-only targets over the grandfather clock targets in Experiment 2 (Chapter 2). Specifically, if oddness is a feature that makes visual search more efficient (Wolfe, 1994; 2007), then detection advantages for body-only targets in Experiment 2 (Chapter 2) can be interpreted in terms of a bias towards their oddness in the scenes, which made search for these targets more efficient (faster). There is no research, other than that presented in this thesis, relating to human bodies and whether the omission of the head makes them appear odd and thus more likely to attract attention in visual scenes. It would therefore be beneficial to explore further the role of oddness using the additional singleton paradigm.

One possible way to further examine whether attention is captured by the oddness of headless bodies is to use other odd control stimuli as distractors and compare the effect their presence has on target search compared to the distractor effect of headless human bodies. For this purpose, non-body distractors could be made odd by removing vital features, e.g., an animal stimulus could be presented with its head removed. Note however, that care in deciding the type of target and distractor stimuli must be taken. This is because evidence from Experiments 8 (Chapter 4), 9 and 10 (Chapter 5) indicated that if target and distractor shared a number of low-level features, the distractor slows search for the target (Wolfe, 1994; 2007). The evidence presented throughout this thesis has suggested that oddness appears to attract attention, at least to some degree. Therefore, it is possible that the presence of odd control distractors will slow the search for a target relative to when they are absent. It is hypothesized that if attention is attracted to odd stimuli, then when the distractor effect of an odd distractor is compared to the distractor effect of a headless body there should be no difference in their respective distractor effects. The experiment outlined here would provide further insight into whether odd in a visual scene attracts attention.

6.7 The Use of Natural Backgrounds to Present Face and Body Stimuli

So far, this discussion has largely been concerned with attentional biases towards human faces, headless bodies and full-bodies. It now returns to the use of the natural backgrounds that these stimuli were presented in. Issues of natural versus artificial backgrounds being a question that this work originally set out to explore. Most previous research concerning face and body perception has used artificial visual displays such as uniform colour background to present stimuli. As such, there were concerns about whether observed attentional advantages for faces (Bindemann & Burton, 2008; Bindemann, et al., 2007; Langton et al., 2008; Ro, Friggel, & Lavie, 2007; Ro et al., 2001; Theeuwes & Van der Stigchel, 2006) and bodies (Ro et al., 2007; Downing et al., 2004; Mack & Rock, 1998) were confined to these impoverished displays (Bindemann et al., 2010). This therefore raised the question of how these previous findings might then be translated to face and body processing in the real world. Taking this into account, the specific aim was to present all stimuli in natural backgrounds to examine whether human faces and bodies have attentional processing advantages that facilitates attention capture by these stimuli.

On the one hand, the findings of Experiments 2, (Chapter 2), 4 (Chapter 3), and 10 (Chapter 5) strongly indicate that human faces have an attentional priority processing over other objects when presented in both natural and artificial (grey) scenes. These findings are broadly congruent with previous suggestions of the preferential processing of faces in artificial scenes (Bindemann & Burton, 2008; Bindemann, et al., 2007; Langton et al., 2008; Ro et al., 2007; Ro et al., 2001; Theeuwes & Van der Stigchel, 2006). However, the current findings also extended previous work in an important way. Specifically, the findings from Experiment 2 (Chapter 2) showed that human faces (face-only targets) had detection advantages over comparable stimuli in natural scenes. Further, Experiment 4 (Chapter 3) showed that regardless of whether they were presented in natural or grey backgrounds, the presence of a human face significantly slowed the search for a target. Experiment 10 (Chapter 5) showed that the human face depicted within a full-human body slowed the search for a target. Thus attention capture by faces appears to be a robust effect, replicable using different paradigms and consistent across different backgrounds of varying complexity. Crucially, this would indicate that attention capture by faces are not confined to artificial backgrounds but also occur in naturalistic backgrounds. As such, this thesis provides a more comprehensive understanding in to the processes by which faces are detected and perceived, under the highly variable viewing conditions encountered in everyday life (Bindemann et al., 2010; Bindemann & Burton, 2009).

On the other hand, evidence of attention capture by the large size of the bodies contrasts directly with conclusions drawn from previous research. Specifically, previous evidence had suggested that distractor bodies capture attention (Downing et al., 2004) and that there is an attentional bias for bodies (Ro et al., 2007). This discrepancy can be explained in the way the stimuli were presented. For example, Downing et al. used uniform colour backgrounds to present the body stimuli, and Ro et al. presented body parts, as part of an array, again in a uniform colour background. Given that all of the stimuli in Downing et al. or Ro et al., were approximately the same size, their findings cannot be explained in terms of attention capture by the size of the bodies. However, considering that Ro et al. used body parts as the body stimuli, it is possible that their results can be explained in terms of a bias towards the oddness of their stimuli. However, without further testing this interpretation is speculation. Indeed it is possible that human bodies do have attentional advantages when presented in uniform colour backgrounds (Downing et al., 2004), or when as part of an array (Ro et al., 2007), but that these advantages do not translate to more realistic contexts, such as the natural backgrounds used in the current thesis.

The findings in this thesis have raised the question of how the current findings and other laboratory research conclusions can then translate to the application of visual search in the more complex real world (Bindemann, 2010b). Specifically, in real life, the dimensions of a computer screen, like in laboratory studies, do not constrict visual information. As such, to what extent can results from visual search using natural backgrounds in the laboratory by extrapolated and generalised to real-world visual search? For example, a number of studies have demonstrated that there is a tendency for viewer to look at the centre of a *scene* and computer *screen* first, suggesting that this is a potential artefact for visual perception studies in laboratory settings that may not be present under real world circumstances (Bindemann, 2010; Bindemann et al., 2010). One possible alternative to using laboratory methods is eye tracking with viewers outside of the laboratory. One way to achieve this would be measure participants eye movements by measuring first eye fixation, speed of fixation and fixation duration to faces and bodies in the real world. In this scenario, participants would be allowed to move around various person present environments wearing eye-tracking glasses. This could either be in staged or naturally occurring scenarios. It would be predicted that if attention is captured by faces then there would be more first fixations, faster responses and longer dwell time for faces when they first arrived in the environment. If the large size of bodies captures attention then it would be expected that there would be equal first fixations, responses and dwell time for bodies and other objects of a similar physical size. This might provide an insight into face and body processing in real-world situations. However, this approach is challenging, given that in a real-world situation it is very difficult to control extraneous variables.

The use of natural backgrounds to present stimuli therefore provides a reasonable compromise between presenting stimuli in artificial backgrounds, which is a completely unrealistic appraisal of visual search, and using eye-tracking methods in the real world where all possible extraneous variables are impossible to control. Extending this, a combination of eye tracking in the laboratory using natural backgrounds to present stimuli may also be a suitable compromise. However, given the lack of research using natural backgrounds to present face and body stimuli, our results reported here and in Bindemann et al. (2010) provides a strong starting point to understanding how faces and bodies are processed under real world circumstances. For a full understanding of face and body detection and perception, future research should continue to present stimuli within more realistic contexts, such as the natural backgrounds like those used in this thesis.

6.8 Contribution of the Current Research to Bindemann et al. (2010)

The starting point of this research was to replicate an original finding from Bindemann et al. (2010), who demonstrated equal detection times for human faces and bodies in natural scenes. This finding suggested that human bodies may have attentional advantages similar to that of faces and as such may have a detection advantage when presented in natural scenes. As such, given the current conclusions, this thesis adds to Bindemann et al.'s (2010) research in several ways. First, Bindemann et al. (2010) did not compare the detection times for the human targets to any other stimuli, as such it was not known whether advantages for human bodies would be held when in comparison to other objects. Extending their findings, in this thesis human faces, bodies and full-bodies were both used as both targets (Experiments 1 – 3) and distractors (4 – 10), in comparison to a variety of other objects. As such, contrary to Bindemann et al. (2010) the present research has rigorously examined and concluded over a number of experiments that bodies do not have detection advantages in natural scenes. Instead, it appears that the bodies being odd and large in size in the scenes facilitates any observed detection advantages. Furthermore, as discussed in detail in section 6.5, the finding of an apparent relationship between target size and detectability suggests that the lack of a similar conclusion in Bindemann et al. may have been the result of the body and full-body targets being partially occluded by other objects in their naturalistic scenes. Finally, as discussed in section 6.7 both the current research and Bindemann et al. (2010) demonstrate the importance of using natural backgrounds to present human targets.

6.9 Limitations of Comparing Results Between Multiple Experiments into the Conclusions

Throughout this thesis, the general conclusions have been inferred from multiple experiments by comparing specific results from one experiment to specific results from another. For example, the conclusion that attention is captured by the relative size of the human bodies rather than the nature of the body stimuli themselves is drawn from evidence across several experiments. Specifically, in Experiments 1 – 3, where the bodies were the targets of search in a visual search paradigm, the larger targets were detected faster than the smaller targets, indicating that target detectability appeared to be related to target size. Further evidence for this came from Experiments 5, 6 and 8 where, in an additional singleton paradigm participants searched for a non-body target in the presence or absence of a body or non-body distractor. It was observed that when the body was the distractor *and* the largest object in the scene, the presence of a human body slowed the visual search for the target, relative to its absence. Seemingly confirming this is in Experiments 7 and 9, where when the human body distractors were size-matched to the targets, they no longer slowed the visual search for the target.

Therefore, the conclusion that attention is captured by the size of the human bodies was drawn from a number of experiments using two different paradigms, one in which the body was the target of search (Experiments 1 – 3) and other (Experiments 5 – 9) in which it was a distractor in the visual search scene. As exemplified here, these experiments were very different from one another, yet were all taken as evidence in favour of attention capture by size of the human bodies. Furthermore, between Experiments 5, 6, 7, 8 and 9, participants were searching for different targets in the scenes, therefore again using different tasks, which may have caused participants to employ different search strategies. It is therefore plausible that the nature of the target stimuli (target/distractor) and the difference in paradigms may have influenced how participants responded.

In the above example, and more throughout the current thesis, there are a number of factors that vary substantially between these experiments. Intrinsically, the different paradigms used may elicit different cognitive processes. Therefore, comparing the specific results from one experiment to specific results of another experiment may appear to provide support for the same hypothesis. However, given the variation in a number of factors between them, to what extent can we, or indeed any researcher using multiple experiments to draw conclusions from, be confident that different experiments are measuring and reflecting the same underlying cognitive process? It therefore cannot be entirely ruled out that the current results may reflect other factors. While I do not believe that this detracts away from the value of the conclusions of this thesis, it is worth acknowledging this as a potential problem in experimental research.

7 Thesis Conclusion

The central aim of this thesis was to examine whether human faces and bodies have an attentional priority processing advantages that allows them to be detected and processed rapidly when presented in natural scenes. This thesis applied two different attentional paradigms address this question. The research carried out here has demonstrated attention capture by human faces in comparison to other stimulus categories. This was shown in both natural and grey scenes; therefore extending previous findings of advantages for faces to more natural contexts. Importantly, it also demonstrates that human bodies do not capture attention in either natural or grey scenes. It is therefore suggested that any apparent detection advantages for human bodies are partly because of an attentional priority bias towards the face and partly because they are relatively bigger targets or distractor objects in a scene. Body processing advantages are also partly explained by the oddness of the body, when it is a headless body. Finally, this research puts forward important avenues of investigation for future research. It recommends that research on priority processing for human faces and bodies take place using natural scenes. Furthermore, in order to develop a full understanding of face and body processing within the real world, it must also carefully consider (even explore) potential confounds such as target/distractor size and various forms of familiarity/unusualness.

8 References

Abrams, R. A., & Christ, S. E. (2006). Motion onset captures attention: A rejoinder to franconeri and simons (2005). *Perception & Psychophysics, 68*(1), 114-117. http://dx.doi.org/10.3758/bf03193661

Andrews, T., J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. *NeuroImage, 17(2),* 890-901. http://dx.doi.org/10.1006/nimg.2002.1243.

Andrews, T. J., & Ewbank, M. P. (2004). Distinct representations for facial identity and changeable aspects of faces in the human temporal lobe. *Neuroimage, 23*(3), 905-913. http://dx.doi.org/10.1016/j.neuroimage.2004.07.060

Arnott, S. R., Pratt, J., Shore, D. I., & Alain, C. (2001). Attentional set modulates visual areas: An event-related potential study of attentional capture. *Cognitive Brain Research, 12*(3), 383-395. http://dx.doi.org/10.1016/j.neuroimage.2004.07.060

Azevedo, E., Squella, S., & Ribeiro-do-Valle, L. (2001). The early facilitatory effect of a peripheral spatially non-informative prime stimulus depends on target stimulus features. *Brazilian Journal of Medical and Biological Research, 34*(6), 803-813. http://dx.doi.org/10.1590/s0100-879x2001000600017.

Bannerman, R. L., Milders, M. & Sahraie, A. (2010). Attentional cueing: Fearful body postures capture attention with saccades. *Journal of Vision, 10*(5), 1-14. http://dx.doi.org/10.1167/10.5.23

Bartels, A., & Zeki, S. (2004). Functional brain mapping during free viewing of natural scenes. *Human Brain Mapping, 21*(2), 75-85. http://dx.doi.org/10.1002/hbm.10153

Barton, J. J., Press, D. Z., Keenan, J. P., & O'Connor, M. (2002). Lesions of the fusiform face area impair perception of facial configuration in prosopagnosia. *Neurology*, *58*, 71-78. http://dx.doi.org/10.1212/wnl.58.1.71

Becker, S. I. (2007). Irrelevant singletons in pop out search: Attentional capture or filtering costs? *Journal of Experimental Psychology: Human Perception and Performance, 33*(4), 764. http://dx.doi.org/10.1037/0096-1523.33.4.764.

Becker, M. W., Pashler, H. & Lubin, J. (2007). Object intrinsic oddities draw early saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 20-30. http://dx.doi.org/10.1037/0096-1523.33.1.20

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B (Methodological),* 289-300. http://dx.doi.org/10.1111/j.1467-9868.2010.00746.x

Berti, A., & Rizzolatti, G. (1992). Visual processing without awareness: Evidence from unilateral neglect. *Journal of Cognitive Neuroscience, 4*(4), 345-351. http://dx.doi.org/10.1162/jocn.1992.4.4.345

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

Bindemann, M., Burton, A. M., Hooge, I. T., Jenkins, R., & de Haan, E. H. (2005a). Faces retain attention. *Psychonomic Bulletin & Review, 12*(6), 1048-1053. http://dx.doi.org/10.3758/bf03206442

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

Bindemann, M., Burton, A. M., Langton, S. R., Schweinberger, S. R., & Doherty, M. J. (2007). The control of attention to faces. *Journal of Vision, 7*(10), 15.1-15.8. http://dx.doi.org/10.1167/7.10.15

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

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision Research, 50*(23), 2577-2587. http://dx.doi.org/10.1016/j.visres.2010.08.016

Birmingham, E., Bischof, W. F., & Kingstone, A. (2008). Social attention and real-world scenes: The roles of action, competition and social content. *The Quarterly Journal of Experimental Psychology, 61*(7), 986-998. http://dx.doi.org/10.1080/17470210701410375

Brandman, T., & Yovel, G. (2010). The body inversion effect is mediated by face-selective, not body-selective, mechanisms. *The Journal of Neuroscience : The*

Official Journal of the Society for Neuroscience, 30(31), 10534-10540. http://dx.doi.org/10.1523/jneurosci.0911-10.2010

Brandman, T., & Yovel, G. (2012). A face inversion effect without a face. *Cognition, 125*(3), 365-372. http://dx.doi.org/10.1016/j.cognition.2012.08.001.

Broadbent, D. E. (1958). *Perception and Communication.* Pergamon press, New York. http://dx.doi.org/10.1037/10037-000

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

Bruder, C. & Ribeiro-do-Valle, L. E. (2009). Influence of separate and mixed experimental designs on reaction times to two simple visual stimuli. *Psychology & Neuroscience, 2(1)*, 3-9. http://dx.doi.org/10.1590/s1983-32882009000100002.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review, 97*(4), 523 – 547. http://dx.doi.org/10.4135/9781483328768.n3

Bundesen, C., Habekost, T., & Kyllingsbæk, S. (2005). A neural theory of visual attention: Bridging cognition and neurophysiology. *Psychological Review, 112*(2), 291 – 328. http://dx.doi.org/10.1037/0033-295x.112.2.291.

Burton, A. et al (2009). Human face detection in natural scenes: Full Research Report ESRC End of Award Report, RES-062-23-0389. Swindon: ESRC.

Carey, S., & Diamond, R. (1977). From piecemeal to configural representation of faces. *Science*, *195*, 312-314. http://dx.doi.org/10.1126/science.831281.

Carey, S. (1992). Becoming a face expert. *Philosophical Transactions of the Royal Society of London — Series B: Biological Sciences 335(1273)*, 95-102. http://dx.doi.org/10.1111/j.1467-9280.2006.01806.x

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

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

Cave, K. R. & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology, 22*, 225-271. http://dx.doi.org/10.1016/0010-0285(90)90017-x.

Chan, L. K. H., & Hayward, W. G. (2009). Feature Integration Theory revisited: Dissociating feature detection and attentional guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 35*(1), 119-132. http://dx.doi.org/10.1037/0096-1523.35.1.119.

Chan, A. W., Peelen, M. V., & Downing, P. E. (2004). The effect of viewpoint on body representation in the extrastriate body area. *Neuroreport, 15*(15), 2407-2410. http://dx.doi.org/10.1097/00001756-200410250-00021.

Chun, M. M., & Wolfe, J. M. (2001). Visual attention. In E. B. Goldstein (Ed.) Blackwell handbook of *Sensation and perception*. Blackwell Publishing, Malden, MA, 272 – 310. http://dx.doi.org/10.1002/9780470753477.

Chun, M. M., & Marois, R. (2002). The dark side of visual attention. *Current Opinion in Neurobiology, 12*(2), 184-189. http://dx.doi.org/10.1016/s0959-4388(02)00309-4.

Cooper, R. M., Law, A. S., & Langton, S. R. (2013). Looking back at the stare-in-the-crowd effect: Staring eyes do not capture attention in visual search. *Journal of Vision, 13*(6), 1-22. http://dx.doi.org/10.1167/13.6.10

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. http://dx.doi.org/10.1167/12.9.822.

Cox, D., Meyers, E., & Sinha, P. (2004). Contextually evoked object-specific responses in human visual cortex. *Science, 304*(5667), 115-117. http://dx.doi.org/10.1126/science.1093110.

Curby, K. M., Glazek, K., & Gauthier, I. (2009). A visual short-term memory advantage for objects of expertise. *Journal of Experimental Psychology: Human Perception & Performance, 35(1),* 94–107. http://dx.doi.org/10.1037/0096-1523.35.1.94

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

Sciences of the United States of America, 100(22), 13105-13110. http://dx.doi.org/10.1073/pnas.1735530100

Damasio, A.R., Damasio, H., and Van Hoesen, G.W. (1982). Prosopagnosia: anatomical basis and behavioural mechanism. *Neurology, 32*, 331-341. http://dx.doi.org/10.1212/wnl.32.4.331.

Deco, G., Pollatos, O., & Zihl, J. (2002). The time course of selective visual attention: Theory and experiments. *Vision Research, 42*(27), 2925-2945. http://dx.doi.org/10.1016/s0042-6989(02)00358-9.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience, 18*(1), 193-222. http://dx.doi.org/10.1146/annurev.neuro.18.1.193.

Devue, C., Laloyaux, C., Feyers, D., Theeuwes, J. & Bredart, S. (2009). Do picture of faces, and which ones, capture attention in the inattentional-blindness paradigm? *Perception*, *38(4),* 552-568. http://dx.doi.org/10.1068/p6049.

Deutsch, J. A., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*(1), 80-90. http://dx.doi.org/10.1057/9781137092861.0004.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*(2), 107-117. http://dx.doi.org/10.1037//0096-3445.115.2.107.

Dijksterhuis, A., & Aarts, H. (2003). On wildebeests and humans: The preferential detection of negative stimuli. *Psychological Science, 14*(1), 14-18. http://dx.doi.org/10.1111/1467-9280.t01-1-01412.

Downing, P. E., Bray, D., Rogers, J. & Childs, C. (2004). Bodies capture attention when nothing is expected. *Cognition*, *93(1)*, B27-38. http://dx.doi.org/10.1016/j.cognition.2003.10.010.

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

Downing, P. E. & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception, *Cognitive Neuroscience*, *2(2-3),* 186-226. http://dx.doi.org/10.1080/17588928.2011.582945.

Driver, J. (2001). A selective review of selective attention research from the past century. *British Journal of Psychology*, *92*(1), 53-78. DOI: 10.1348/000712601162103

Duncan, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review, 87*(3), 272-300. http://dx.doi.org/10.1037//0033-295x.87.3.272.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96*(3), 433 – 458. http://dx.doi.org/10.3758/bf03193880.

Engell, A. D., & Haxby, J. V. (2007). Facial expression and gaze-direction in human superior temporal sulcus. *Neuropsychologia, 45*(14), 3234-3241. http://dx.doi.org/10.1016/j.neuropsychologia.2007.06.022.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a non search task. *Perception & Psychophysics, 16*(1), 143-149. http://dx.doi.org/10.3758/bf03203267.

Eriksen, C. W., & Hoffman, J. E. (1972). Temporal and spatial characteristics of selective encoding from visual displays. *Perception & Psychophysics, 12*(2), 201-204. http://dx.doi.org/10.3758/bf03212870.

Eriksen. E. & Spencer. T. (1969) Rate of information processing in visual perception: Some results and methodological considerations, *Journal of Experimental Psychology Monograph,* 79, 1 – 16. http://dx.doi.org/10.1037/h0026873

Farah, M. J., Klein, K. L., & Levinson, K. L. (1995). Face perception and within-category discrimination in prosopagnosia. *Neuropsychologia, 33,* 661-674. http://dx.doi.org/10.1016/0028-3932(95)00002-k.

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. http://dx.doi.org/10.1167/3.9.819.

Farah, M. J., Wilson, K. D., Drain, H. M. & Tanaka, J. W. (1998). What is “special” about face perception? *Psychological Review, 105*(3), 482-498. http://dx.doi.org/10.1093/acprof:oso/9780199230136.001.0001.

Fenske, M. J., & Eastwood, J. D. (2003). Modulation of focused attention by faces expressing emotion: Evidence from flanker tasks. *Emotion, 3*(4), 327-343. http://dx.doi.org/10.1037/1528-3542.3.4.327

Findlay, J. M. (1997). Saccade target selection during visual search. *Vision Research, 37*(5), 617-631. http://dx.doi.org/10.1016/s0042-6989(96)00218-0.

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

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance, 18*(4), 1030-1044. http://dx.doi.org/10.1037//0096-1523.18.4.1030.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1993). Contingent attentional capture: A reply to Yantis (1993). *Journal of Experimental Psychology: Human Perception and Performance, 19(3),* 682-685. http://dx.doi.org/10.1037/0096-1523.19.3.682.

Folk, C. L., Remington, R. W., & Wright, J. H. (1994). The structure of attentional control: Contingent attentional capture by apparent motion, abrupt onset, and colour. *Journal of Experimental Psychology: Human Perception and Performance, 20*(2), 317-329. http://dx.doi.org/10.1037/0096-1523.20.2.317.

Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General, 130*(4), 681-700. http://dx.doi.org/10.1037//0096-3445.130.4.681.

Franconeri, S. L., & Simons, D. J. (2003). Moving and looming stimuli capture attention. *Perception & Psychophysics, 65*(7), 999-1010. http://dx.doi.org/10.3758/bf03194829.

French, R. M., Mareschal, D., Mermillod, M., & Quinn, P. C. (2004). The role of bottom-up processing in perceptual categorization by 3-to 4-month-old infants: Simulations and data. *Journal of Experimental Psychology: General, 133*(3), 382-397. http://dx.doi.org/10.1037/0096-3445.133.3.382.

Garrido, L., Duchaine, B., & Nakayama, K. (2008). Face detection in normal and prosopagnosic individuals. *Journal of Neuropsychology, 2*(1), 119-140. http://dx.doi.org/10.1348/174866407x246843.

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience, 3*(2), 191-197. http://dx.doi.org/10.1038/72140.

Gauthier, I., & Tarr, M. J. (1997). Becoming a “Greeble” expert: Exploring mechanisms for face recognition. *Vision Research, 37*(12), 1673-1682. http://dx.doi.org/10.1016/s0042-6989(96)00286-6.

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform face area increases with expertise in recognizing novel objects. *Nature Neuroscience, 2*(6), 568-573. doi:10.1038/9224

Gellatly, A. R. H. & Cole, G. G. (2000). Accuracy of target detection in new-object and old-object displays. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 889–899. http://dx.doi.org/10.1037/0096-1523.26.3.889.

Gibson, B. S., & Kelsey, E. M. (1998). Stimulus-driven attentional capture is contingent on attentional set for display wide visual features. *Journal of Experimental Psychology: Human Perception and Performance, 24*(3), 699-706. http://dx.doi.org/10.1037/0096-1523.24.3.699.

Gilchrist, I. D., & Proske, H. (2006). Anti-saccades away from faces: Evidence for an influence of high-level visual processes on saccade programming. *Experimental Brain Research*, *173*, 708–712. http://dx.doi.org/10.1007/s00221-006-0455-1.

Goren, C. C., Sarty, M., & Wu, P. Y. (1975). Visual following and pattern discrimination of face-like stimuli by newborn infants. *Pediatrics, 56*(4), 544-549.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annual Review of Neuroscience, 27*, 649-677. http://dx.doi.org/10.1146/annurev.neuro.27.070203.144220

Hansen, C. H., & Hansen, R. D. (1988). Finding the face in the crowd: An anger superiority effect. *Journal of Personality and Social Psychology, 54*(6), 917 – 927. http://dx.doi.org/10.1037//0022-3514.54.6.917.

Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., and Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, *293*, 2425–2430. http://dx.doi.org/10.1126/science.1063736.

Haxby, J. V., Ungerleider, L. G., Clark, V. P., Schouten, J. L., Hoffman, E. A., & Martin, A. (1999). The effect of face inversion on activity in human neural systems for face and object perception. *Neuron, 22*(1), 189-199. http://dx.doi.org/10.1016/s0896-6273(00)80690-x.

Haxby, J. V., Grady, C. L., Horwitz, B., Ungerleider, L. G., Mishkin, M., Carson, R. E., et al. (1991). Dissociation of object and spatial visual processing pathways in human extrastriate cortex. *Proceedings of the National Academy of Sciences of the United States of America. 88*(5), 1621-1625. http://dx.doi.org/10.1016/b978-0-08-042004-2.50028-9.

Heilman, K. M., Watson, R. T., & Valenstein, E. (1993). Neglect and related disorders. *Clinical Neuropsychology, 20(4)*, 279-336. http://dx.doi.org/10.1055/s-2008-1041551.

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

Hershler, O., & Hochstein, S. (2009). The importance of being expert: Top-down attentional control in visual search with photographs. *Attention, Perception, & Psychophysics, 71*(7), 1478-1486. http://dx.doi.org/10.3758/app.71.7.1478.

Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. *Journal of Vision, 10*(10), 1 – 14. http://dx.doi.org/10.1167/10.10.21.

Hodzic, A., Kaas, A., Muckli, L., Stirn, A., & Singer, W. (2009). Distinct cortical networks for the detection and identification of human body. *Neuroimage, 45*(4), 1264-1271. http://dx.doi.org/10.1016/j.neuroimage.2009.01.027.

Honk, J. V., Tuiten, A., de Haan, E., vann de Hout, M., & Stam, H. (2001). Attentional biases for angry faces: Relationships to trait anger and anxiety. *Cognition & Emotion, 15*(3), 279-297. http://dx.doi.org/10.1080/0269993004200222.

Huang, L., Song, Y., Li, J., Zhen, Z., Yang, Z. & Liu, J. (2014). Individual differences in cortical face selectivity predict behavioural performance in face recognition. *Frontiers in Human Neuroscience*, *18*(483), 1-10. http://dx.doi.org/10.3389/fnhum.2014.00483

Hwang, A. D., Higgins, E. C., & Pomplun, M. (2009). A model of top-down attentional control during visual search in complex scenes. *Journal of Vision, 9*(5), 1-18. http://dx.doi.org/10.1167/9.5.25.

Ishai, A. (2008). Let's face it: It's a cortical network. *Neuroimage, 40*(2), 415-419. http://dx.doi.org/10.1016/j.neuroimage.2007.10.040.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10), 1489-1506. http://dx.doi.org/10.1016/s0042-6989(99)00163-7.

Jenkins, R., Burton, A. M., & Ellis, A. W. (2002). Long-term effects of covert face recognition. *Cognition, 86*(2), B43-B52. http://dx.doi.org/10.1016/s0010-0277(02)00172-5.

Jiang, Y., & He, S. (2006). Cortical responses to invisible faces: Dissociating subsystems for facial-information processing. *Current Biology, 16*(20), 2023-2029. http://dx.doi.org/10.1016/j.cub.2006.08.084.

Jiang, Y., Costello, P., & He, S. (2007). Processing of invisible stimuli: Advantage of upright faces and recognizable words in overcoming interocular suppression. *Psychological Science, 18*(4), 349-355. http://dx.doi.org/10.1111/j.1467-9280.2007.01902.x.

Kahneman, D. (1973). *Attention and effort.* Englewood Cliffs, NJ; Prentice-Hall

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

Kanwisher, N. & Yovel, G. (2009). Cortical Specialization for Face Perception in Humans. In Cacioppo, J.T. & Berntson, G.G. (Eds.), *Handbook of Neuroscience for the Behavioral Sciences*. J. Wiley and Sons.

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal*

Society of London.Series B, Biological Sciences, 361(1476), 2109-2128. http://dx.doi.org/10.1098/rstb.2006.1934

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

Kessler, E., Walls, S. A., & Ghuman, A. S. (2013). Bodies adapt orientation-independent face representations. *Frontiers in Psychology, 4* (413), 1-8. http://dx.doi.org/10.3389/fpsyg.2013.00413

Klein, R. (1988). Inhibitory tagging system facilitates visual search. *Nature, 334*(6181), 430-431. http://dx.doi.org/10.1038/334430a0.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return is a foraging facilitator in visual search. *Psychological Science, 10*(4), 346-352. http://dx.doi.org/10.1111/1467-9280.00166.

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. http://dx.doi.org/10.1068/p230095.

Lambert, A., Wells, I., & Kean, M. (2003). Do isoluminant colour changes capture attention? *Perception & Psychophysics, 65*(4), 495-507.). Do isoluminant colour changes capture attention? http://dx.doi.org/10.3758/bf03194578.

Lamy, D., Amunts, L., & Bar-Haim, Y. (2008). Emotional priming of pop out in visual search. *Emotion, 8*(2), 151-161. http://dx.doi.org/10.1037/1528-3542.8.2.151.

Lamy, D., & Tsal, Y. (2001). On the status of location in visual attention. *European Journal of Cognitive Psychology, 13*(3), 305-342. http://dx.doi.org/10.1080/09541440042000016.

Langton, S. R., Law, A. S., Burton, A. M., & Schweinberger, S. R. (2008). Attention capture by faces. *Cognition, 107*(1), 330-342. http://dx.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. http://dx.doi.org/10.1037//0096-1523.21.3.451

Lavie, N. (2000). Selective attention and cognitive control: dissociating attentional functions through different types of load. In S. Monsell & J. Driver (Eds.). *Attention and performance XVIII*, pp. 175-194. Cambridge, Massachusetts: MIT press.

Lavie, N., Lin, Z., Zokaei, N., & Thoma, V. (2009). The role of perceptual load in object recognition. *Journal of Experimental Psychology: Human Perception and Performance, 35*(5), 1346-1358. http://dx.doi.org/10.1037/a0016454.

Lavie, N., Ro, T., & Russell, C. (2003). The role of perceptual load in processing distractor faces. *Psychological Science, 14*(5), 510-515. http://dx.doi.org/10.1111/1467-9280.03453.

Leber, A. B., & Egeth, H. E. (2006). It's under control: Top-down search strategies can override attentional capture. *Psychonomic Bulletin & Review, 13*(1), 132-138. http://dx.doi.org/10.3758/bf03193824.

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. http://dx.doi.org/10.1002/ima.10040

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

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

Liu, J., Harris, A., & Kanwisher, N. (2010). Perception of face parts and face configurations: An fMRI study. *Journal of Cognitive Neuroscience, 22*(1), 203-211. http://dx.doi.org/10.1162/jocn.2009.21203.

Mack, A., Pappas, Z., Silverman, M., & Gay, R. (2002). What we see: Inattention and the capture of attention by meaning. *Consciousness and Cognition, 11*(4), 488-506. http://dx.doi.org/10.1016/s1053-8100(02)00028-4.

Mack, A., & Rock, I. (1998). Inattentional blindness: Perception without attention. *Visual Attention, 8*, 55-76.

Malach, R., Reppas, J. B., Benson, R. R., Kwong, K. K., Jiang, H., Kennedy, W. A., et al. (1995). Object-related activity revealed by functional magnetic resonance imaging in human occipital cortex. *Proceedings of the National Academy of Sciences of the United States of America, 92*(18), 8135-8139. http://dx.doi.org/10.1073/pnas.92.18.8135.

McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-specific processing in the human fusiform gyrus. *Journal of Cognitive Neuroscience, 9*(5), 605-610. http://dx.doi.org/10.1162/jocn.1997.9.5.605.

McKone, E., Kanwisher, N., & Duchaine, B. C. (2007). Can generic expertise explain special processing for faces? *Trends in Cognitive Sciences, 11*(1), 8-15. http://dx.doi.org/10.1016/j.tics.2006.11.002.

McLeod, P., Driver, J., & Crisp, J. (1988). Visual search for a conjunction of movement and form is parallel. *Nature.* 332(6160), 154-155. http://dx.doi.org/10.1038/332154a0.

Meadows, J. (1974). Disturbed perception of colours associated with localized cerebral lesions. *Brain. 97*(4), 615-632. http://dx.doi.org/10.1093/brain/97.1.615.

Minnebusch, D. A., Suchan, B., & Daum, I. (2009). Losing your head: Behavioural and electrophysiological effects of body inversion. *Journal of Cognitive Neuroscience, 21*(5), 865-874. http://dx.doi.org/10.1162/jocn.2009.21074.

Mitchell, K. J., Livosky, M. & Mather, M. (1999). The weapon focus effect revisited: The role of novelty. *Legal and Criminological Psychology*, *3*(2), 287-303. http://dx.doi.org/10.1111/j.2044-8333.1998.tb00367.x

Mogg, K., & Bradley, B. P. (1999). Orienting of attention to threatening facial expressions presented under conditions of restricted awareness. *Cognition & Emotion, 13*(6), 713-740. http://dx.doi.org/10.1080/026999399379050.

Moore, C. D., Cohen, M. X., & Ranganath, C. (2006). Neural mechanisms of expert skills in visual working memory. *The Journal of Neuroscience : The Official*

Journal of the Society for Neuroscience, 26(43), 11187-11196. http://dx.doi.org/10.1523/jneurosci.1873-06.2006.

Moran, R., Zehetleitner, M., Liesefeld, H. R., Müller, H. J., & Usher, M. (2015). Serial vs. parallel models of attention in visual search: Accounting for benchmark RT-distributions. *Psychonomic Bulletin & Review,* 1-16. http://dx.doi.org/10.3758/s13423-015-0978-1

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology, 11*(1), 56-60. http://dx.doi.org/10.1080/17470215908416289.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to cousineau (2005). *Reason, 4*(2), 61-64.

Morris, J. P., Pelphrey, K. A., & McCarthy, G. (2007). Face processing without awareness in the right fusiform gyrus. *Neuropsychologia, 45*(13), 3087-3091. http://dx.doi.org/10.1016/j.neuropsychologia.2007.05.020.

Moscovitch, M., Winocur, G., & Behrmann, M. (1997). What is special about face recognition? nineteen experiments on a person with visual object agnosia and dyslexia but normal face recognition. *Cognitive Neuroscience, Journal of, 9*(5), 555-604. http://dx.doi.org/10.1162/jocn.1997.9.5.555.

Müller, H. J., & Krummenacher, J. (2006). Visual search and selective attention. *Visual Cognition, 14*(4-8), 389-410. http://dx.doi.org/10.1080/13506280500527676.

Murray, J. E., Rhodes, G., & Schuchinsky, M. (2003). When is a face not a face? In M. A. Peterson & G. Rhodes (Eds.), Perception of faces, objects, and scenes (pp.

75-91). New York: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780195313659.003.0004.

Nakayama, K., & Silverman, G. H. (1986). Serial and parallel processing of visual feature conjunctions. *Nature, 320*(6059), 264-265. http://dx.doi.org/10.1038/320264a0.

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

Neisser, U. (1979). The control of information pickup in selective looking. In A. D. Pick (Ed.), *Perception and its development: A tribute to Eleanor J. Gibson* (pp. 201-219). Hillsdale, NJ: Lawrence Erlbaum. http://dx.doi.org/10.2307/1422379.

Nestor, A., Vettel, J. M., & Tarr, M. J. (2008). Task-specific codes for face recognition: How they shape the neural representation of features for detection and individuation. *PloS One, 3*(12), e3978. http://dx.doi.org/10.1371/journal.pone.0003978.

Nichols, D. F., Betts, L. R., & Wilson, H. R. (2010). Decoding of faces and face components in face-sensitive human visual cortex. *Frontiers in Psychology, 1*(28), 1-13. http://dx.doi.org/10.3389/fpsyg.2010.00028.

Nickerson, R. S. (1965). Short-term memory for complex meaningful visual configurations: A demonstration of capacity. *Canadian Journal of Psychology/Revue Canadienne De Psychologie, 19*(2), 155-160. http://dx.doi.org/10.1037/h0082899.

Norman, D. A. (1968). Toward a theory of memory and attention. *Psychological Review, 75*(6), 522-536. http://dx.doi.org/10.1037/h0026699.

Nothdurft, H. (1993). Faces and facial expressions do not pop out. *Perception-London-, 22*, 1287-1287. http://dx.doi.org/10.1068/p221287.

Nothdurft, H. (2000). Salience from feature contrast: Additivity across dimensions. *Vision Research, 40*(10), 1183-1201. http://dx.doi.org/10.1016/s0042-6989(00)00031-6.

Öhman, A., Flykt, A., & Esteves, F. (2001). Emotion drives attention: Detecting the snake in the grass. *Journal of Experimental Psychology: General, 130*(3), 466 – 478. http://dx.doi.org/10.1037//0096-3445.130.3.466.

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review, 108*(3), 483-522. http://dx.doi.org/10.1037//0033-295x.108.3.483.

Olk, B. & Garay-Vado, A. M. (2011) Attention to faces: effects of face inversion. *Vision Research*, *51(14)*, 1659-1666. http://dx.doi.org/10.1167/10.7.125.

Op de Beeck, H. P., Baker, C. I., DiCarlo, J. J., & Kanwisher, N. G. (2006). Discrimination training alters object representations in human extrastriate cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 26*(50), 13025-13036. http://dx.doi.org/10.1523/jneurosci.2481-06.2006.

Palermo, R., & Rhodes, G. (2007). Are you always on my mind? A review of how face perception and attention interact. *Neuropsychologia, 45*(1), 75-92. http://dx.doi.org/10.1016/j.neuropsychologia.2006.04.025.

Palmer, J. & Mclean, J. (1995). *Imperfect, unlimited-capacity, parallel search yields* *large set-size effects*. Paper presented at the annual meeting of the Society for Mathematical Psychology. Irvine, CA. Retrieved from. http://faculty.washington.edu/jpalmer/files/palmer_mclean1995talk.pdf.

Palmer, J., Verghese, P., & Pavel, M. (2000). The psychophysics of visual search. Vision Research, 40(10-12), 1227–1268. http://dx.doi.org/10.1016/ S0042-6989(99)00244-8

Pashler, H. E. (1998). *The psychology of attention.* Cambridge, MA: MIT Press.

Pashler, H. E. (1984). Evidence against late selection: Stimulus quality effects in previewed displays. *Journal of Experimental Psychology: Human Perception and Performance, 10*(3), 429-448. http://dx.doi.org/10.1037//0096-1523.10.3.429.

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

Peirce, J. W. (2007). PsychoPy - psychophysics software in python. *Journal of Neuroscience Methods, 162(1-2)*, 8-13. http://dx.doi.org/10.1016/j.jneumeth.2006.11.017.

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., et al. (1997). A specific neural substrate for perceiving facial expressions of disgust. *Nature, 389*(6650), 495-498. http://dx.doi.org/10.1038/39051.

Pitcher, D., Walsh, V., & Duchaine, B. (2011). The role of the occipital face area in the cortical face perception network. *Experimental Brain Research, 209*(4), 481-493. http://dx.doi.org/10.1007/s00221-011-2579-1.

Pitcher, D., Walsh, V., Yovel, G., & Duchaine, B. (2007). TMS evidence for the involvement of the right occipital face area in early face processing. *Current Biology, 17*(18), 1568-1573. http://dx.doi.org/10.1016/j.cub.2007.07.063

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*(1), 3-25. http://dx.doi.org/10.1093/acprof:oso/9780199791217.003.0003.

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

Posner, M. I., & Snyder, C. R. R. (1975a). Facilitation and inhibition in the processing of signals. *Attention and Performance.* In P. M. A. Rabbit & S. Dornic (Eds.), *Attention and Performance V* (669 – 682). New York: Academic Press.

Posner, M. I., & Snyder, C. R. (1975b). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition* (pp. 55–85). Hillsdale, NJ: Erlbaum.

Proulx, M. J. (2010). Size matters: Large objects capture attention in visual search. *PloS One, 5*(12), e15293. http://dx.doi.org/10.1371/journal.pone.0015293.

Proulx, M. J., & Egeth, H. E. (2008). Biased competition and visual search: The role of luminance and size contrast. *Psychological Research, 72*(1), 106-113. http://dx.doi.org/10.1007/s00426-006-0077-z.

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

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 16*(16), 5205-5215. http://dx.doi.org/10.1016/s1053-8119(96)80364-8.

Puce, A., Allison, T., Bentin, S., Gore, J.C., McCarthy, G., 1998. Temporal cortex activation in humans viewing eye and mouth movements. *Journal of Neuroscience. 18,* 2188–2199.

Purcell, D. G., & Stewart, A. L. (1986). The face-detection effect. *Bulletin of the Psychonomic Society, 24*(2), 118-120.

Purcell, D. G., Stewart, A. L., & Skov, R. B. (1996). It takes a confounded face to pop out of a crowd. *Perception, 25(9)*, 1091-1120. http://dx.doi.org/10.1068/p251091.

Quinlan, P. T. (2003). Visual feature integration theory: Past, present, and future. *Psychological Bulletin, 129*(5), 643-673. http://dx.doi.org/10.1037/0033-2909.129.5.643.

Ratcliff R. A theory of memory retrieval (1978). *Psychological Revie*w, 85, 59–108.

Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 849-860. http://dx.doi.org/10.1037//0096-1523.18.3.849.

Reed, C. L., Stone, V. E., Grubb, J. D., & McGoldrick, J. E. (2006). Turning holistic processing upside down: Part and whole body postures. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 73-87. http://dx.doi.org/10.1037/0096-1523.32.1.73.

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science, 14*(4), 302-308. http://dx.doi.org/10.1111/1467-9280.14431.

Rees, G., Kreiman, G., Koch., C. (2002). Neural correlates of consciousness in humans. *Nature Reviews Neuroscience, 3*, 261-270. http://dx.doi.org/10.1038/nrn783.

Remington, R., Johnston, J. C., & Yantis, S. (1992). Involuntary attentional capture by abrupt onsets. *Perception & Psychophysics, 51,* 279-290. http://dx.doi.org/10.3758/bf03212254.

Remington, R. W., Folk, C. L., & McLean, J. P. (2001). Contingent attentional capture or delayed allocation of attention? *Perception & Psychophysics, 63*(2), 298-307. http://dx.doi.org/10.3758/bf03194470.

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

Rhodes, G., Brake, S., & Atkinson, A. P. (1993). What's lost in inverted faces? *Cognition, 47*(1), 25-57. http://dx.doi.org/10.1016/0010-0277(93)90061-y.

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

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

Robbins, R. A., & Coltheart, M. (2012). The effects of inversion and familiarity on face versus body cues to person recognition. *Journal of Experimental Psychology.Human Perception and Performance, 38*(5), 1098-1104. http://dx.doi.org/10.1037/a0028584.

Rossion, B. (2008). Picture-plane inversion leads to qualitative changes of face perception. *Acta Psychologica, 128*(2), 274-289. http://dx.doi.org/10.1016/j.actpsy.2008.02.003.

Rousselet, G. A., Mace, M. J., & Fabre-Thorpe, M. (2003). Is it an animal? Is it a human face? Fast processing in upright and inverted natural scenes. *Journal of Vision, 3*(6), 440-455. http://dx.doi.org/10.1167/3.6.5.

Sagi, D., & Julesz, B. (1985). "Where" and "what" in vision. *Science. 228*(4704), 1217-1219. http://dx.doi.org/10.1126/science.4001937.

Said, C. P., Moore, C. D., Engell, A. D., Todorov, A., & Haxby, J. V. (2010). Distributed representations of dynamic facial expressions in the superior temporal sulcus. *Journal of Vision, 10*(5), 1-19. http://dx.doi.org/10.1167/10.5.11.

Schwarzlose, R. F., Baker, C. I., & Kanwisher, N. (2005). Separate face and body selectivity on the fusiform gyrus. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 25*(47), 11055-11059. http://dx.doi.org/10.1523/jneurosci.2621-05.2005.

Scialfa, C. T., & Thomas, D. M. (1994). Age differences in same-different judgments as a function of multidimensional similarity. *Journal of Gerontology, 49*(4), P173-8. http://dx.doi.org/10.1093/geronj/49.4.p173.

Sergent, J. (1984). An investigation into component and configural processes underlying face perception. *British Journal of Psychology, 75,* 221-242. http://dx.doi.org/10.1111/j.2044-8295.1984.tb01895.x

Sergent, J., Ohta, S., & MacDonald, B. (1992). Functional neuroanatomy of face and object processing. A positron emission tomography study. *Brain : A Journal of Neurology, 115 Pt 1*, 15-36. http://dx.doi.org/10.1093/brain/115.1.15.

Sergent J., Signoret J. L. (1992). Varieties of functional deficits in prosopagnosia. *Cerebral Cortex, 2,* 375-388. http://dx.doi.org/10.1093/cercor/2.5.375.

Shaw, J. I. & Skolnick, P. (1994). Sex differences, weapon focus and eyewitness reliability. *Journal of Applied Psychology, 134*(4), 413-420. http://dx.doi.org/10.1080/00224545.1994.9712191.

Shapiro, K. L. (1994). The attentional blink: The brain's "eyeblink". *Current Directions in Psychological Science (Wiley-Blackwell), 3*(3), 86-89. http://dx.doi.org/10.1111/1467-8721.ep10770436.

Shelley-Tremblay, J., & Mack, A. (1999). Metacontrast masking and attention. *Psychological Science, 10*(6), 508-515. http://dx.doi.org/10.1111/1467-9280.00197.

Shen, J. & Reingold, E. L. (2001). Visual search asymmetry: The influence of stimulus familiarity and low-level features. *Perception & Psychophysics*, *63*(3), 465-575. http://dx.doi.org/10.3758/bf03194413.

Slaughter, V., Stone, V. E., & Reed, C. (2004). Perception of faces and bodies: Similar or different? *Current Directions in Psychological Scien*ce, *13*, 219–223. http://dx.doi.org/10.1111/j.0963-7214.2004.00312.x.

Snodgrass, J. G., & Townsend, J. T. (1980). Comparing parallel and serial models: Theory and implementation. *Journal of Experimental Psychology: Human Perception Performance, 6,* 330 – 354. doi: http://dx.doi.org/10.1037/0096-1523.6.2.330

Spiridon, M., Fischl, B., & Kanwisher, N. (2006). Location and spatial profile of category- specific regions in human extrastriate cortex. *Human Brain Mapping, 27*(1), 77-89. http://dx.doi.org/10.1002/hbm.20169.

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. http://dx.doi.org/10.1016/j.cognition.2012.06.005.

Steinman, S. (1987). Serial and parallel search in pattern vision? *Perception, 16*(3), 389-398. http://dx.doi.org/10.1068/p160389.

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*(3736), 652-654. http://dx.doi.org/10.1126/science.153.3736.652.

Styles, E. A. (1997). *The Psychology of Attention.* Psychology Press. http://dx.doi.org/10.4324/9780203016435.

Tanaka, J. W., & Farah, M. (1993). Parts and wholes in face recognition. *Quarterly Journal of Experimental Psychology, 46*, 225-245. http://dx.doi.org/10.1037/e665402011-519.

Tanaka, J.,W. & Gauthier, I. (1997). Expertise in object and face recognition. In: Goldstone, R., L., Schyns, P., G & Medin, D.L., editors. *Psychology of Learning and Motivation*, vol. 36. San Diego, CA: Academic Press, 83 – 125. http://dx.doi.org/10.1016/s0079-7421(08)60282-0

Taylor, J. C., Wiggett, A. J., & Downing, P. E. (2007). Functional MRI analysis of body and body part representations in the extrastriate and fusiform body areas. *Journal of Neurophysiology, 98*(3), 1626-1633. http://dx.doi.org/10.1152/jn.00012.2007.

Taubert, J., Apthorp, D., Aageten-Murphy & Alais, D. (2001). The role of the holistic processing in face perception: evidence from the face inversion effect. *Vision Research, 51(11),* 1273-1278. http://dx.doi.org/10.1016/j.visres.2011.04.002.

Techentin, C., & Voyer, D. (2007). Congruency, attentional set, and laterality effects with emotional words. *Neuropsychology, 21*(5), 646-655. http://dx.doi.org/10.1037/0894-4105.21.5.646.

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics, 49*(1), 83-90. http://dx.doi.org/10.3758/bf03211619.

Theeuwes, J. (1992). Perceptual selectivity for colour and form. *Perception & Psychophysics, 51*(6), 599-606. http://dx.doi.org/10.3758/bf03211656.

Theeuwes, J. (1994). Endogenous and exogenous control of visual selection. *Perception-London, 23*(4), 429-440. http://dx.doi.org/10.1068/p230429.

Theeuwes, J. (1996). Visual search at intersections: An eye-movement analysis. *Vision in Vehicles, 5*, 125-134.

Theeuwes, J., Kramer, A. F., Hahn, S., Irwin, D. E. & Zelinsky, G. J. (1999). Influence of attentional capture on occulomotor control. *Journal of Experimental Psychology: Human Perception and Performance, 25*(6), 1595-1608. http://dx.doi.org/10.1037/e536982012-682.

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition, 13*(6), 657-665. http://dx.doi.org/10.1080/13506280500410949.

Thornton, T. L., & Gilden, D. L. (2007). Parallel and serial processes in visual search. *Psychological Review*, 114(1), 71–103. http://dx.doi.org/10.1037/ 0033-295X.114.1.71

Townsend, J. T. (1972). Some results concerning the identifiability of parallel and serial processes. British Journal of Mathematical and Statistical Psychology, 25, 168–199. http://dx.doi.org/10.1111/j.2044-8317.1972.tb00490.x

Townsend, J. T. (1976). Serial and within-stage independent parallel model equivalence on the minimum completion time. Journal of Mathematical Psychology, 14, 219–238. http://dx.doi.org/10.1016/0022-2496(76)90003-1

Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedledee but they can (and should) be distinguished.

Psychological Science, 1, 46–54. http://dx.doi.org/10.1111/j.1467- 9280.1990.tb00067.x

Tipper, S. P. (1985). The negative priming effect: Inhibitory priming by ignored objects. *The Quarterly Journal of Experimental Psychology, 37*(4), 571-590. http://dx.doi.org/10.1080/14640748508400920.

Treisman, A., Sykes, M., & Gelade, G. (1977). Selective attention and stimulus integration. In S. Dornic (Ed.), *Attention and performance VI* (pp. 333-361). Hillsdale: Lawrence Erlbaum.

Treisman, A. (1982). Perceptual grouping and attention in visual search for features and for objects. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 194. http://dx.doi.org/10.1037/0096-1523.8.2.194.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing, 31*(2), 156-177. http://dx.doi.org/10.1016/s0734-189x(85)80004-9.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American, 255*(5), 114-125. http://dx.doi.org/10.1038/scientificamerican1186-114b.

Treisman, A. M. (1969). Strategies and models of selective attention. *Psychological Review, 76*(3), 282-299. http://dx.doi.org/10.1037/h0027242.

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

Treisman, A., & Geffen, G. (1967). Selective attention: Perception or response? *The Quarterly Journal of Experimental Psychology, 19*(1), 1-17. http://dx.doi.org/10.1080/14640746708400062.

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*(1), 15-48.

Treisman, A., & Sato, S. (1990). Conjunction search revisited. *Journal of Experimental Psychology: Human Perception and Performance, 16*(3), 459-478. http://dx.doi.org/10.1037//0096-1523.16.3.459.

Treisman, A., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General, 114*(3), 285-310. http://dx.doi.org/10.1037//0096-3445.114.3.285.

Tsotsos, J. K. (1990). Analyzing vision at the complexity level. *Behavioural and Brain Sciences, 13*(03), 423-445. http://dx.doi.org/10.1017/s0140525x00079577.

Tsuchiya, N. Koch, C. (2005). Continuous flash suppression reduces negative afterimages. *Nature Neuroscience*, *8*, 1096–1101. http://dx.doi.org/10.1038/nn1500.

van Koningsbruggen, M. G., Peelen, M. V., & Downing, P. E. (2013). A causal role for the extrastriate body area in detecting people in real-world scenes. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 33*(16), 7003-7010. http://dx.doi.org/10.1523/jneurosci.2853-12.2013.

VanRullen, R. (2006). On second glance: Still no high-level pop out effect for faces. *Vision Research, 46*(18), 3017-3027. http://dx.doi.org/10.1016/j.visres.2005.07.009.

Vavassis, A., & Von Grunau, M. W. (2007). Complex backgrounds delay low-load visual search. *Spatial Vision, 20*(5), 467-488. http://dx.doi.org/10.1163/156856807781503631.

Verghese, P. (2001). Visual search and attention: A signal detection theory approach. Neuron, 31(4), 523–535. http://dx.doi.org/10.1016/s0896-6273(01)00392-0

Vuilleumier, P. (2000). Faces call for attention: Evidence from patients with visual extinction. *Neuropsychologia, 38*(5), 693-700. http://dx.doi.org/10.1016/s0028-3932(99)00107-4

Vuilleumier, P., Armony, J., Clarke, K., Husain, M., Driver, J., & Dolan, R. J. (2002). Neural response to emotional faces with and without awareness: Event-related fMRI in a parietal patient with visual extinction and spatial neglect. *Neuropsychologia, 40*(12), 2156-2166. http://dx.doi.org/10.1016/s0028-3932(02)00045-3.

Wang, Q., Cavanagh, P., & Green, M. (1994). Familiarity and pop out in visual search. *Perception & Psychophysics, 56*(5), 495-500. http://dx.doi.org/10.3758/bf03206946.

Ward, R. & McClelland, J. L. (1989). Conjunctive search for one and two identical targets. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 664-672. http://dx.doi.org/10.1037/0096-1523.15.4.664.

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage, 52*(4), 1559-1573. http://dx.doi.org/10.1016/j.neuroimage.2010.04.262.

Winston, J. S., Henson, R. N., Fine-Goulden, M. R., & Dolan, R. J. (2004). fMRI-adaptation reveals dissociable neural representations of identity and expression in face perception. *Journal of Neurophysiology, 92*(3), 1830-1839. http://dx.doi.org/10.1152/jn.00155.2004.

Wolfe, J. M. (1994). Guided search 2.0 a revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202-238. http://dx.doi.org/10.3758/bf03200774.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33-39. http://dx.doi.org/10.1111/1467-9280.00006.

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics, 63*(3), 381-389. http://dx.doi.org/10.3758/bf03194406.

Wolfe J. M. (2007). Guided Search 4.0: Current progress with a model of visual search. In Gray W. (Ed.), Integrated models of cognitive systems (pp. 99–119). New York: Oxford. http://dx.doi.org/10.1093/acprof:oso/9780195189193.003.0008

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

Wolfe, J. M., & Horowitz, T. S. (2004). Opinion: What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*(6), 495-501. http://dx.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. http://dx.doi.org/10.1016/s0042-6989(02)00388-7.

Yantis, S. (1993). Stimulus-driven attentional capture. *Current Directions in Psychological Science. 2(5)*,156-161. http://dx.doi.org/10.1111/1467-8721.ep10768973.

Yantis, S., & Johnson, D. N. (1990). Mechanisms of attentional priority. *Journal of Experimental Psychology: Human Perception and Performance, 16*(4), 812 – 825. http://dx.doi.org/10.1037/0096-1523.16.4.812.

Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: Voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16*(1), 121-134. http://dx.doi.org/10.1037/0096-1523.16.1.121.

Yantis, S., & Egeth, H. E. (1999). On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology.Human Perception and Performance, 25*(3), 661-676. http://dx.doi.org/10.1037/0096-1523.25.3.661.

Yantis, S. (1996). Attentional capture in vision. In A. F. Kramer, M. G. H. Coles & G. D. Logan (Eds.), (pp. 45-76). Washington, DC, US: American Psychological Association. http://dx.doi.org/10.1037/10187-002

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

Young, A. W., Ellis, A. W., Flude, B. M., McWeeny, K. H., & Hay, D. C. (1986). Face–name interference. *Journal of Experimental Psychology: Human Perception and Performance, 12*(4), 466-475. http://dx.doi.org/10.1037/0096-1523.12.4.466.

Yovel, G., Pelc, T., & Lubetzky, I. (2010). It's all in your head: Why is the body inversion effect abolished for headless bodies? *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 759-767. http://dx.doi.org/10.1037/a0017451.

Yue, X., Tjan, B. S., & Biederman, I. (2006). What makes faces special? *Vision Research, 46*(22), 3802-3811. http://dx.doi.org/10.1016/j.visres.2006.06.017.

Xu, Y. (2005). Revisiting the role of the fusiform face area in visual expertise. *Cerebral Cortex*, 15(8), 1234–1242. http://dx.doi.org/10.1093/cercor/bhi006.

9 Appendices.

9.1 Appendix 1. Benjamini and Hochberg Corrected Pairwise Comparisons for Experiment 3.

Paired comparisons also revealed that human face-only targets (*M* = 537.65) were detected significantly faster than clock face-only targets (*M* = 551.21, *d* = 0.40, $p < 0.05$). Clock full-body targets (*M* = 517.48) were detected significantly faster than human face-only targets (*M* = 537.68; $p < 0.05$, *d* = 0.37). Human body-only targets (*M* = 527.06) were detected significantly faster than clock face-only (*M* = 551.21) and clock body-only targets (*M* = 540.26; $p < 0.05$, *d* = 0.37). Additionally, human full-body targets (*M* = 516.03) were detected significantly faster than clock face-only (*M* = 551.21, *d* = 0.46) and body-only targets (*M* = 540.24; $p < 0.05$, *d* = 0.41). No other comparison reached significance (all *p*'s > 0.05).

9.2 Appendix 2. Benjamini and Hochberg Corrected Pairwise Comparisons for Experiment 4

Paired comparisons also revealed that RTs in natural target-absent/distractor-absent scenes (*M* = 629.52) were significantly slower than RTs n both grey target-present/distractor-absent (*M* = 571.33, $p < 0.05$, $d = 0.79$) and grey target-present/distractor-present (*M* = 593.12, $p < 0.05$, $d = 0.47$).

RTs in natural target-absent/distractor-present scenes (*M* = 673.67) were significantly slower than RTs in both natural target-present/distractor-absent (*M* = 608.45, $d = 1.30$) and RTs in natural target-present/distractor-present scenes (*M* = 628.00, $d = 1.09$). RTs in natural target-absent/distractor-present scenes (*M* = 673.67) were significantly slower than RTs in both grey target-present/distractor-absent scenes (*M* = 671.33, $p < 0.05$, $d = 1.27$) and grey target-present/distractor-present scenes (*M* = 593.12, $p < 0.05$, $d = 0.93$). RTs in natural target-present/distractor-absent scenes (*M* = 608.45) were significantly slower than RTs in grey target-absent/distractor-absent scenes (*M* = 750.02, $p < 0.05$, $d = 0.82$) and RTs in grey target-absent/distractor-present scenes (*M* = 649.45, $p < 0.05$, $d = 0.75$). RTs for natural target-present/distractor-present scenes (*M* = 628.00) were significantly faster than RTs in grey target-absent/distractor-absent scenes (M = 750.02, $p < 0.05$, $d = 0.71$). RTs were significantly slower in grey target-absent/distractor-absent scenes (*M* = 750.02) than RTs in both grey target-present/distractor-absent trials (*M* = 571.33, $p < 0.05$, $d = 1.05$) and grey target-present/distractor-present scenes (*M* = 593.12, $p < 0.05$, $d = 0.83$). RTs for grey target-absent/distractor-present scenes (*M* = 649.45) were significantly slower than RTs in both grey target-present/distractor-absent scenes (*M* = 571.33, $p < 0.05$, $d = 1.92$) and grey target-present/distractor-present scenes (*M* = 593.12, $p < 0.05$, $d = 0.98$). RTs in natural target-present/distractor-present scenes (*M* = 628.00) were significantly slower than RTs in grey target-present/distractor-absent scenes (*M* = 571.33, $p < 0.05$, $d = 0.88$). No other comparison reached significance, $p > 0.05$.

9.3 Appendix 3. Benjamini and Hochberg Corrected Pairwise Comparisons for reaction time data in Experiment 5.

RTs for natural target-absent/distractor-absent scene (*M* = 609.75) were significantly slower for natural target-present/distractor-absent scenes (*M* = 573.52, $p < 0.05$, *d* = 0.61), grey target-present/distractor-absent (*M* = 552.65, $p < 0.05$, *d* = 0.89) and grey target-present/distractor-present scenes (*M* = 565.71, $p < 0.05$, *d* = 0.72)

RTs for natural target-absent/distractor-present scenes (*M* = 630.63) were significantly slower than RTs for natural target-present/distractor-absent (*M* = 573.52, $p < 0.05$, *d* = 1.12), natural target-present/distractor-present (*M* = 600.64, $p < 0.05$, *d* = 0.51), grey target-present/distractor-absent (*M* = 552.65, $p < 0.05$, *d* = 1.39) and grey target-present/distractor-present (*M* = 565.71, $p < 0.05$, *d* = 1.25) scenes.

RTs for natural target-present/distractor-absent scenes (*M* = 573.52, $p < 0.05$) were significantly faster than RTs for both grey target-absent/distractor-absent (*M* = 680.39, *d* = 0.81) and grey target-absent/distractor-present (*M* =616.24, $p < 0.05$, *d* = 0.59). RTs for natural target-present/distractor-present scenes (*M* = 600.64) were significantly faster than RTs for grey target-absent/distractor-absent scenes (*M* = 680.39, $p < 0.05$, *d* = 0.79) and significant slower than RTs for grey target-present/distractor-absent scenes (*M* = 552.65, $p < 0.05$, *d* = 0.88). RTs for grey target-absent/distractor-absent scenes (*M* = 680.39) were significantly slower than RTs in both grey target-present/distractor-absent scenes (*M* = 552.65, $p < 0.05$, *d* = 1.09) and target-present/distractor-present scenes (*M* = 564.71, $p < 0.05$, *d* = 1.08). RTs for grey target-absent/distractor-present scenes (*M* = 616.24) were significantly slower than RTs in both grey target-present/distractor-absent (*M* = 552.65, $p < 0.05$, *d* = 1.30) and grey target-present/distractor-present scenes (*M* = 565.71, $p < 0.05$, *d* = 1.25)

9.4 Appendix 4. Benjamini and Hochberg Corrected Pairwise Comparisons for accuracy data in Experiment 5.

There were significantly fewer errors made in grey target-absent/distractor-absent scenes ($M = 2.05\%$) than natural target-present/distractor-present scenes ($M = 3.77\%$, $p < 0.05$, $d = 0.38$). There were significantly more errors made in grey target-absent/distractor-present scenes ($M = 4.56$v) than in grey target-present/distractor-absent ($M = 2.31\%$, $p < 0.05$, $d = 0.57$), grey target-present/distractor-present ($M = 1.92\%$, $p < 0.05$, $d = 0.57$), natural target-absent/distractor-absent scenes ($M = 1.19\%$, $p < 0.05$, $d = 0.75$), and natural target-present/distractor-absent scenes ($M = 2.18$, $p < 0.05$, $d = 0.41$). There were significantly more errors made in grey target-present/distractor-absent ($M = 2.31\%$) scenes than natural target-absent/distractor-absent scenes ($\mathrm{M} = 1.19\%$, $p < 0.05$, $d = 0.46$. There were significantly fewer errors made in natural target-absent/distractor-absent ($M = 1.19\%$) scenes than in natural target-present/distractor-absent scenes ($M = 2.18\%$, $p < 0.05$, $d = 0.32$) and natural target-present/distractor-present scenes ($M = 3.77\%$, $p < 0.05$, $d = 0.84$). There were significantly more errors made in natural target-present/distractor-present scenes ($M = 3.77\%$) than in natural target-present/distractor-absent scenes ($M = 2.18\%$, $p < 0.05$, $d = 0.38$) and in natural target-absent/distractor-present scenes ($M = 1.79\%$, $p < 0.05$, $d = 3.91$).

9.5 Appendix 5. Benjamini and Hochberg Corrected Pairwise Comparisons for reaction time data in Experiment 6.

RTs in natural target-absent/distractor-absent scenes (M = 662.02) were significant faster than RTs in grey target-target scenes (M = 575.52, $p < 0.05$, d = 0.72). RTs in natural distractor-distractor scenes (M = 691.10) were significantly slower than RTs in natural target-target scenes (M = 591.83, $p < 0.05$, d = 0.82), grey target-target scenes (M = 575.52, $p < 0.05$, d = 1.03) and natural distractor-distractor scenes (M = 609.13, $p < 0.05$, d = 0.57).

RTs in natural target-target scenes (M = 591.83) were significantly faster than RTs in both grey target-absent/distractor-absent scenes (M = 752.77, $p < 0.05$, d = 1.20) and grey distractor-distractor scenes (M = 644.07, $p < 0.05$, d = 0.81). RTs for natural target-distractor scenes (M = 633.50) were significantly faster than RTs in both grey target-absent/distractor-absent scenes (M = 757.77, $p < 0.05$, d = 0.71) and significantly slower than RTs in grey target-target scenes (M = 575.52, $p < 0.05$, d = 0.80). RTs for grey target-absent/distractor-absent (M = 752.77) were significantly slower than RTs in both grey target-target (M = 575.52, $p < 0.05$, d = 1.32) and RTs in grey target-distractor (M = 609.13, $p < 0.05$, d = 1.78). RTs in grey distractor-distractor scenes (M = 644.07) significantly slower than RTs in both grey target-target scenes (M = 575.52, $p < 0.05$, d = 1.91) and grey target-distractor scenes (M = 609.13, $p < 0.05$, d = 0.64).

9.6 Appendix 6. Benjamini and Hochberg Corrected Pairwise Comparisons for accuracy data in Experiment 6.

There were significantly more errors made in grey target-absent/distractor-absent scenes (M = 3.15%) than target-target scenes (M = 0.93%, $p < 0.01$, d = 0.53). There were also significantly more errors in grey distractor-distractor (M = 3.98%) scenes than natural target-absent/distractor absent scenes (M = 1.48%, $p < 0.05$, d = -0.63). There were significantly more errors in grey target-distractor scenes (M = 4.91%) than in natural target-absent/distractor-absent scenes (M = 1.48%, $p < 0.05$, d = -0.64). There were significantly more errors in natural target-distractor scenes (M = 6.39%) than natural target-absent/distractor-absent scenes (M = 1.48%, $p < 0.01$, d = -0.77). There were significantly more errors made in grey distractor-distractor scenes (M = 3.98%) than grey target-target scenes (M = 0.93%, $p < 0.05$, d = 0.70). There were significantly more errors made in grey distractor-distractor scenes (M = 3.98%) than natural target-target scenes (M = 1.67%, $p < 0.05$, d = 0.47). There were significantly more errors made in natural target-distractor scenes (M = 6.39%) than natural distractor-distractor scenes (M = 2.59%, $p < 0.05$, d = -0.53). There were significantly more errors made in natural target-distractor scenes (M = 6.39%) than in grey target-target scenes (M = 0.93%, $p < 0.05$, d = -1.26). There were also significantly more errors made in grey target-distractor scenes (M = 4.91%) than natural target-target scenes (M = 1.67%, $p < 0.05$, d = -0.71).

9.7 Appendix 7. Benjamini and Hochberg Corrected Pairwise Comparisons for reaction time data in Experiment 8.

RTs for target-absent/distractor-absent scenes (*M* = 684.75) were significantly slower than RTs for target-present/distractor scenes (*M* = 617, $p < 0.05$, $d = 0.77$), target-present/body-distractor scenes (*M* = 640.51, $p < 0.05$, $d = 0.55$), and target-present/lamp-distractor scenes (*M* = 649.98, $p < 0.05$, $d = 0.45$). RTs for target-absent/body-distractor scenes (*M* = 674.13) were significantly slower than RTs in target-present/distractor-absent scenes (*M* = 617.28, $p < 0.05$, $d = 0.83$), target-present/body-distractor scenes (*M* = 640.51, $p < 0.05$, $d = 0.62$), and target-present/lamp-distractor scenes (*M* = 649.98 $p < 0.05$, $d = 0.47$). RTs for target-absent/lamp-distractor (*M* = 712.63) were significantly slower than RTs in target-present/distractor-absent scenes (*M* = 617.26, $p < 0.05$, $d = 1.21$), target-present/body-distractor scenes (*M* = 640.51, $p < 0.05$, $d = 1.26$), and target-present/lamp-distractor scenes (*M* = 649.98, $p < 0.05$, $d = 1.34$).

9.8 Appendix 8. Benjamini and Hochberg Corrected Pairwise Comparisons for reaction time data in Experiment 10.

RTs for target-absent/distractor-absent scenes (*M* = 624.90) were significantly slower than RTs for target-present/distractor-absent scenes (*M* = 587.23, $p < 0.05$, $d = 0.57$). RTs for target-absent/body-only scenes (*M* = 626.64) were significantly slower than RTs for target-present/distractor-absent scenes (*M* = 587.23, $p < 0.05$, $d = 0.70$), and target-present/lamp-distractor scenes (*M* = 610.89, $p < 0.05$, $d = 0.44$). RTs for target-absent/full-body scenes (*M* = 648.70) were significantly slower than RTs for target-present/distractor-absent scenes (*M* = 587.23, $p < 0.05$, $d = 0.84$), target-present/body-only scenes (*M* = 613.04, $p < 0.05$, $d = 0.52$), target-present/full-body scenes (*M* = 619.75, $p < 0.05$, $d = 0.49$), and target-present/lamp-distractor scenes (*M* = 610.89, $p < 0.05$, $d = 0.61$). RTs for target-absent/lamp-distractor scenes (*M* = 661.75) were significantly slower than RTs for target-present/distractor-absent scenes (*M* = 587.23, $p < 0.05$, $d = 1.36$), target-present/body-only scenes (*M* = 613.04, $p < 0.05$, $d = 0.96$), target-present/full-body scenes (*M* = 619.75, $p < 0.05$, $d = 0.93$) and target-present/lamp-distractor scenes (*M* = 610.89, $p < 0.05$, $d = 1.09$).