Target Template Guidance of Eye Movements During Real-World Search

George Law Malcolm

Doctor of Philosophy
Department of Psychology
University of Edinburgh
2010

Table of Contents

Abstract

Humans must regularly locate task-relevant objects when interacting with the world around them. Previous research has identified different types of information that the visual system can use to help locate objects in real-world scenes, including low-level image features and scene context. However, previous research using object arrays suggest that there may be another type of information that can guide real-world search: target knowledge. When a participant knows what a target looks like they generate and store a visual representation, or template, of it. This template then facilitates the search process. A complete understanding of real-world search needs to identify how a target template guides search through scenes.

Three experiments in Chapter 2 confirmed that a target template facilitates real-world search. By using an eye-tracker target knowledge was found to facilitate both scanning and verification behaviours during search, but not the search initiation process. Within the scanning epoch a target template facilitated gaze directing and shortened fixation durations. These results suggest that target knowledge affects both the activation map, which selects which regions of the scene to fixate, and the evaluation process that compares a fixated object to the internal representation of the target.

With the exact behaviours that a target template facilitates now identified, Chapter 3 investigated the role that target colour played in template-guided search. Colour is one of the more interesting target features as it has been shown to be preferred by the visual system over other features when guiding search through object arrays. Two real-world search experiments in Chapter 3 found that colour information had its strongest effect on the gaze directing process, suggesting that the visual system relies heavily on colour information when searching for target-similar regions in the scene percept. Although colour was found to facilitate the evaluation process too, both when rejecting a fixated object as a distracter and accepting it as the target, this behaviour was found to be influenced comparatively less. This suggests that the two main search behaviours – gaze directing and region evaluation – rely on different

sets of template features. The gaze directing process relies heavily on colour information, but knowledge of other target features will further facilitate the evaluation process.

Chapter 4 investigated how target knowledge combined with other types of information to guide search. This is particularly relevant in real-world search where several sources of guidance information are simultaneously available. A single experiment investigated how target knowledge and scene context combined to facilitate search. Both information types were found to facilitate scanning and verification behaviours. During the scanning epoch both facilitated the eye guidance and object evaluation processes. When both information sources were available to the visual system simultaneously, each search behaviour was facilitated additively. This suggests that the visual system processes target template and scene context information independently.

Collectively, the results indicate not only the manner in which a target template facilitates real-world search but also updates our understanding of real-world search and the visual system. These results can help increase the accuracy of future real-world search models by specifying the manner in which our visual system utilises target template information, which target features are predominantly relied upon and how target knowledge combines with other types of guidance information.

Acknowledgements

I would like to thank John Henderson and Antje Nuthmann for their guidance, support and patience during my PhD study, as well as members of the Visual Cognition Lab for their invaluable feedback over the last three years.

I would like to thank my wife, Katharine Malcolm, for supporting me across two continents and for her bravery in marrying a student!

Finally, I would like to thank my parents, Drs. Neil and Joyce Malcolm, for their support and encouragement; who encouraged my dreams of pursuing academia at a higher level and instilled in me an interest in vision science from a very young age.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(George Law Malcolm)

Chapter 1: Introduction

As humans we constantly search our environment for task related objects. We look for the coffee machine in the morning, our favourite shirt in the wardrobe, road signs in an unfamiliar neighbourhood, etc. Locating objects is a prerequisite for interacting with the world around us. We are very adept at such real-world search tasks; however, the necessary behaviours and cognitive processing involved are deceptively complex.

Due to the nature of the retina, we can only appreciate about 2° of our visual field with high acuity. This clarity is produced by light projecting onto the fovea, a centrally-located region of the retina with the highest cone density (Palmer, 1999). If we want to perceive an object in full visual detail we must direct the fovea towards it. The fovea must be relatively stable, or fixated, during this time. Due to the small dimensions of the fovea and the comparatively large size of the visual field outside of it, deciding where to fixate next requires selecting a location in the low-resolution periphery. Once a region is selected for fixation, the eyes move in a short ballistic movement, known as a saccade. A fundamental goal in the study of real-world search is to understand how the visual system guides these eye movements.

Despite their frequency (3-4 per second; Rayner, 1998) eye movements are not directed randomly. Instead, eye movements are guided by the visual system integrating low-resolution information in the visual periphery with our knowledge about the current task and environment (Buswell, 1935; Castelhano, Mack, & Henderson, 2009; Yarbus, 1967), constrained by oculomotor regularities (Tatler, 2007; Tatler & Vincent, 2008, 2009; Unema, Pannasch, Joos, & Velichkovsky, 2005). Within this framework several factors determine where the eyes are guided during scene perception. For instance, systematic tendencies in the form of oculomotor behavioural biases (Tatler & Vincent, 2008, 2009; Unema et al., 2005) and a central fixation bias (Tatler, 2007; Tseng, Carmi, Cameron, Munoz, & Itti, 2009) can predict human fixation data with a surprising amount of accuracy, even in the absence of an incoming scene percept. Similarly, eye movements can be guided by memory for a target object's location in a scene (Brockmole, Castelhano, & Henderson, 2006; Brockmole & Henderson, 2006a, 2006b; Hollingworth, 2009).

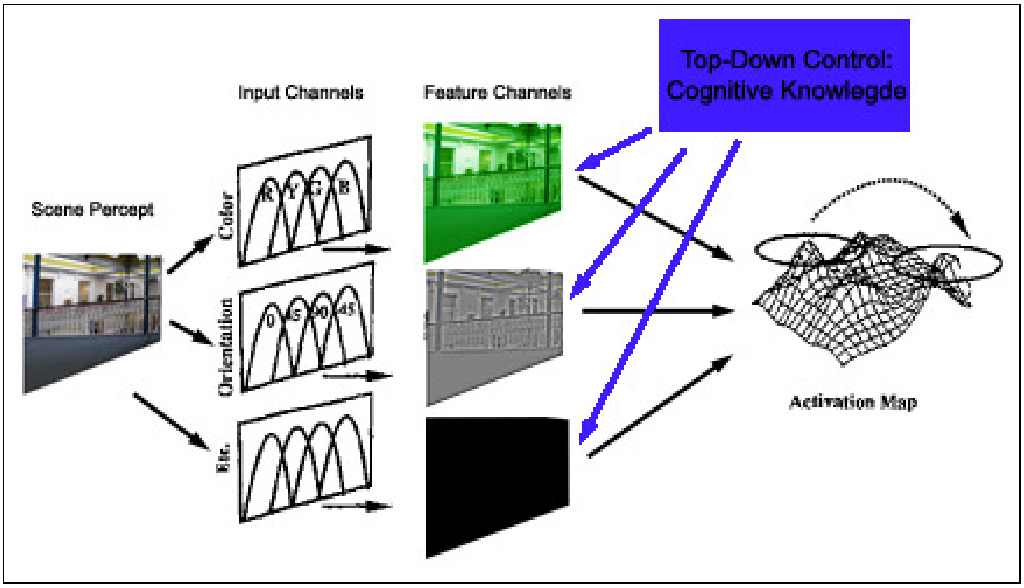

Figure 1 The basic set up of the two-stage feature/activation map model. A scene percept enters the visual system via the retina. The features in the image are separated and analysed in different channels forming topographical maps. Regions within the maps corresponding to task-relevant areas in the visual field receive higher activation. The feature maps are then summed to form a pre-attentive activation map, the highest point of which is selected for focused attention. (Note, parts of this image come from Wolfe, 1994).

Research in the present thesis focuses on another aspect of eye movement guidance in real-world scene images: the processing of information in a novel scene percept. When a novel scene image appears in a display, the human visual system processes the percept so as to extract the necessary information to guide eye movements. This is accomplished in two stages. First the scene percept's basic features are processed in parallel across the visual field, with each feature or dimension organised into respective feature maps (Figure 1). Feature maps are differentially weighted retinotopic maps, with regions corresponding to areas in the visual field. Second, output from feature maps then combine to form an activation map[1]: a topographical, pre-attentive map that indicates where eye movements should be directed, but does not disclose the visual properties at the selected location. Regions in the visual field that the visual system has determined to be task-related receive higher activation,

[1] Also known as a saliency map.

while less interesting regions receive lower activation. The region with highest activation is then selected for fixation. Although there are different theories concerning how the visual system chooses to weight the information represented in the activation map (e.g., Henderson, Malcolm, & Schandl, 2009; Itti & Koch, 2000; Torralba, Oliva, Castelhano, & Henderson, 2006; Wolfe, 1994a), most search models use an interpretation of this feature/activation map structure.

Different types of information within a novel scene percept have been proposed to influence the activation map (Figure 2), with these information types falling into two main categories: image features and cognitive knowledge (Henderson, 2003). When the visual system uses image features it selects regions of contrast in the image – usually on the basis of colour, luminance, and intensity – to guide eye movements (Itti & Koch, 2000, 2001; Koch & Ullman, 1985; Parkhurst, Law, & Niebur, 2002). Under this hypothesis the visual system selects the scene regions with highest contrast to fixate; if the selected region does not contain the target, the eyes are directed to the region with the next highest contrast, and so forth. This is a purely bottom-up form of guidance that acts independently of the current task (but see Navalpakkam & Itti, 2005).

This form of guidance has gained popularity for two main reasons. First, empirical evidence indicates that during scene perception tasks, fixations tend to fall on a non-random selection of features (Mannan, Ruddock, & Wooding, 1996, 1997). More specifically, observers have been found to actively fixate scene regions with higher than average contrast and edge content (Tatler, Baddeley, & Gilchrist, 2005). Secondly, image feature information, as opposed to cognitive knowledge, can be applied to computational models (Itti & Koch, 2000; Itti, Koch, & Niebur, 1998; Parkhurst et al., 2002).

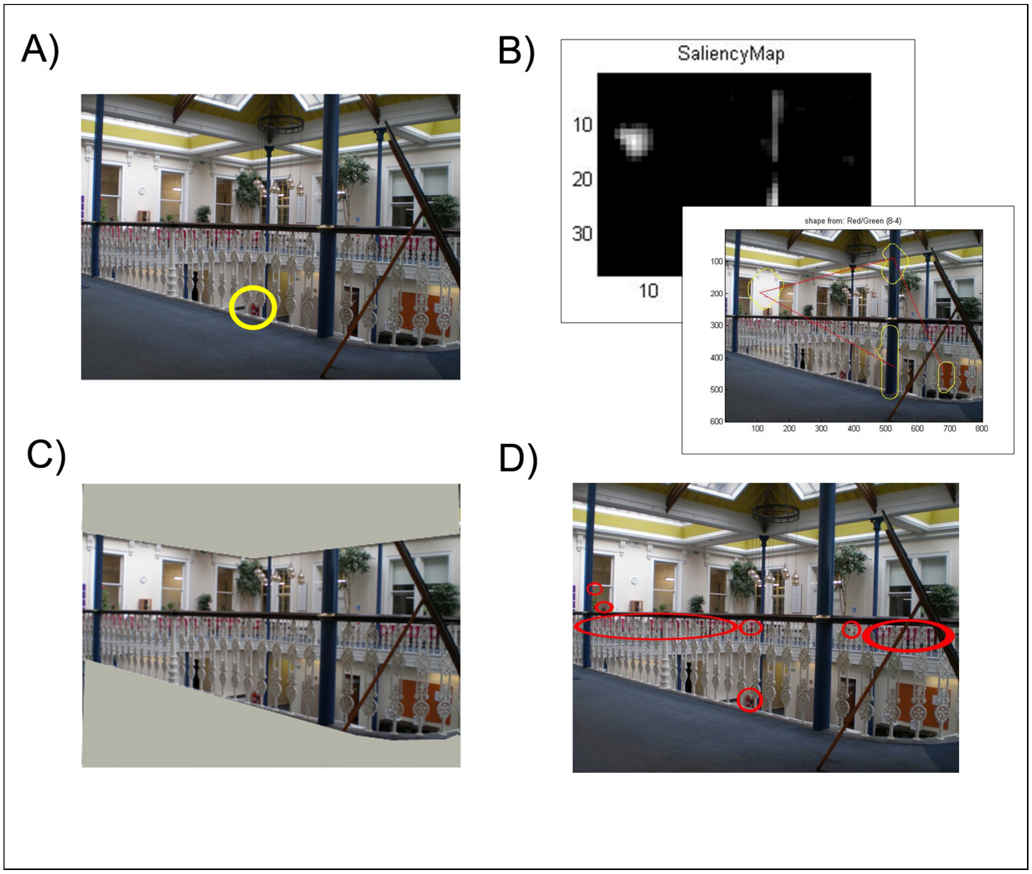

Figure 2 Different types of information are available for guiding eye movements in novel, real-world images. A) The target object in this indoor picture is the fire extinguisher (circled). B) The first type of guidance information is image features: regions of contrast in terms of orientation, luminance and colour are combined to form a saliency map. The visual system then proceeds to select regions to fixate from highest to lowest salience (this image and scanpath were generated using the Itti & Koch (2000) algorithm). C) Next is guidance by scene context: the scene type can be identified and probable target locations selected. The fire extinguisher is unlikely to be positioned on the floor or ceiling so those areas can be ignored. In contrast the wall regions receive higher activations. Eye movements are then directed toward the walls. D) Finally, there is target template guidance: the visual system directs the eyes toward target-similar features in the scene percept (e.g., red regions, shiny regions, etc).

However, there is growing evidence that the studies claiming that image features guide eye movements are confounded. The fixation content analysis may be confounded by the visual system's tendency to fixate objects, which contain more of the features in question than empty spaces (Buswell, 1935; Castelhano et al., 2009; Yarbus, 1967). Similarly, it has been demonstrated that regions selected for fixation are not only correlated with a particular feature content but also semantic informativeness (Henderson, Brockmole, Castelhano, & Mack, 2007). Furthermore, though image feature-based computational models have had some success in predicting human scan patterns, they tend to fail in visual search tasks involving real-world scenes (Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Einhäuser, Rutishauser, & Koch, 2008; Einhäuser et al., 2007; Foulsham & Underwood, 2007; Henderson et al., 2009; Tatler et al., 2005; Tatler & Vincent, 2009; Turano, Geruschat, & Baker, 2003; Zelinsky, 2008; Zelinsky, Zhang, Yu, Chen, & Samaras, 2006). Even the high intersubject agreement of the first few fixations – which was claimed to be a result of the visual system selecting low-level salient regions to fixate (Carmi & Itti, 2006; Mannan, Ruddock, & Wooding, 1995; Parkhurst et al., 2002) – has since been shown to be an artefact, and instead thought to derive from common high-level knowledge strategies (Tatler et al., 2005). For these reasons, it can not unequivocally be claimed that the visual system uses image features to guide eye movements.

The reason image features have minimal influence on eye movements is, most likely, due to their independence from the task goal: the information provided by image features remains constant regardless of task. In contrast, the placement of eye movements is heavily task dependent (Castelhano et al., 2009; Foulsham & Underwood, 2007; Land & Hayhoe, 2001; Land, Mennie, & Rusted, 1999; Malcolm, Lanyon, Fugard, & Barton, 2008; Rothkopf, Ballard, & Hayhoe, 2007; Yarbus, 1967).

As opposed to image features, when the visual system uses cognitive knowledge to guide eye movements it combines the schema of the scene with the task goal to direct eye movements. There are two general forms of cognitive knowledge in novel scene percepts that can guide eye movements. The first is scene context. When the visual system uses scene context to guide eye movements it utilises high-level knowledge to identify the type of scene, usually within the space of a single fixation (Castelhano & Henderson, 2007; Intraub, 1981; Joubert, Rousselet, Fize, & Fabre-Thorpe, 2007; Oliva & Schyns, 2000; Potter, 1975; Potter & Levy, 1969), and then selects global regions within the scene with a high probability of containing the target object (Castelhano & Henderson, 2007; Eckstein, Drescher, & Shimozaki, 2006; Ehinger et al., 2009; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006;

Torralba et al., 2006; see also Chapter 4). For example, when looking for a clock, and the scene is recognised as an office, the scene's context would suggest that the target should be located high on the walls. Corresponding regions on the visual system's activation map receive higher activation while regions corresponding to unlikely locations (e.g., the floor or ceiling) would receive less activation.

From a computational standpoint, models that incorporate some form of scene context information find a relatively high correlation between their generated fixation positions and those of humans (Ehinger et al., 2009; Torralba et al., 2006), unlike pure image feature models.

Target Template Guidance

The visual system can guide search through real-world scenes by using either of the above types of guidance information. Previous research into object-array search, however, suggests that there may be a second form of cognitive knowledge that could guide eye movements through novel scene images: a target template. A target template is the visual system's representation of the target object. If you are looking for the clock in an office, you would create and store a representation of the clock, the features of which will be used to guide search. This guidance occurs by the visual system using the stored representation to bias feature map activity so that regions with properties similar to those of the target receive greater activation. If the sought after clock were red, the visual system would increase the activation of red items in the colour feature map. This would, in turn, increase activity in the respective regions of the activation map, improving those regions' chances of being fixated. Areas in the scene without similar features will receive less activation. In order to fully understand how the visual system guides real-world search, target template guidance has to be demonstrated in scenes and its contributions to real-world search recognised.

Interest in target template guidance grew with the development of the *Guided Search Model* (Wolfe, 1994a; Wolfe, Cave, & Franzel, 1989). The *Guided Search Model* refined popular two-stage model search theories such as *Feature Integration Theory* (Treisman & Gelade, 1980; Treisman, Sykes, & Gelade, 1977) and *Texton Theory* (Julesz, 1984). These early search theories were based on empirical studies using visual arrays, and proposed that immediately after display onset features are registered in parallel across the visual field. If the target differs from distracters by a single feature (colour, for example) it would be detected almost immediately, regardless of set size (see Figure 3). This is sometimes known as a "pop-out" effect. If the target is not detected immediately, usually a result of the target differing from distracters by a conjunction of features, the visual system resorts to the second-stage of processing: serial scanning of the elements in the visual display. This stage is necessary if a combination of features distinguishes the target from distracters. Focused attention is needed to bind features into an object. The focus of attention necessary for binding objects must be directed around the environment, generally overtly in the form of eye movements to inspect separate display elements (see the sequential attention model, Henderson, 1992; or the pre-motor theory of attention, Rizzolatti, Riggio, Dascola, & Umiltá, 1987). Once features are attended and bound into an object (e.g., a red horizontal bar) a target/distracter decision can be made.

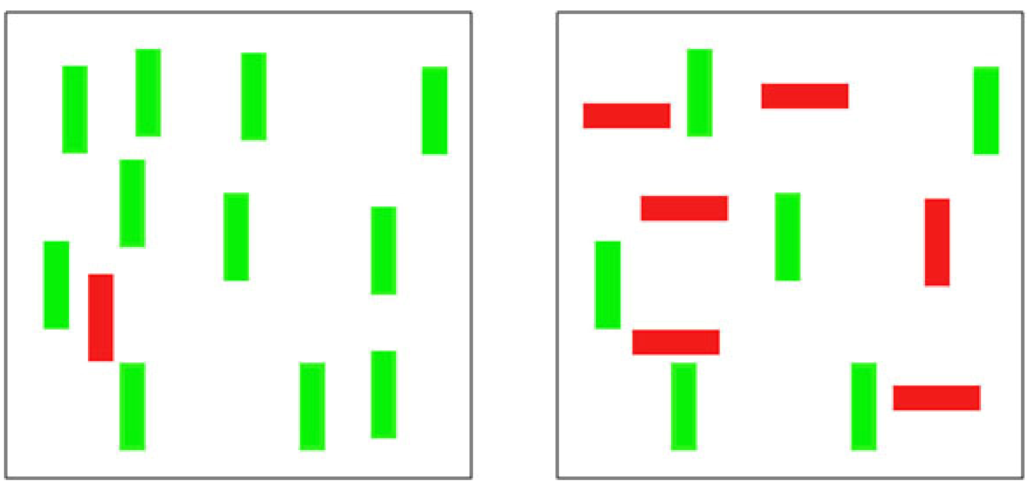

Figure 3 Example of displays that led to the generation of two-stage search models. In both displays there are 12 elements. The task is to find the red, vertical bar. In the left image target detection is immediate since the target object differs from the distracters by one feature. When the visual system processes the display in parallel, the region in the colour feature map corresponding to the target will provide the greatest signal, and the target will seem to "pop-out". In the right image, search will take longer as the target now differs from distracters in one of two possible ways, adding noise to the visual system. The visual system must now go through the scene in a slower manner, allocating attention to individual or nearby groups of objects. This allocation of attention will bind the fixated element's features into an object at which point an accept/reject decision can be made.

Wolfe noted that previous two-stage models did not utilise information about the target's appearance to guide search in the serial stage. This implies that there is no top-down facilitation of search when the target's visual features are known. Thus, if the parallel search process does not reveal a target element, the visual system would treat every stimulus in the display as a possible target. This would mean that the ensuing serial search would be very inefficient, selecting items to focus attention on in a random manner, regardless of whether the viewer knew the target's properties (but see Tatler, 2007; Wolfe, O'Neill, & Bennett, 1998; for examples of search biases without top-down or bottom-up influence).

Instead, Wolfe suggested that serial attention could be guided by matching the target template's visual features with the features gathered during the parallel processing phase, thus separating potential targets from obvious distracters. In other words, information from the parallel processing stage could be used to guide the serial search process. If the target were a red vertical bar among green vertical and red horizontal bars, parallel processing could not detect target presence. However, parallel processing could differentiate red from green items and vertical from horizontal items, and since none of the green or horizontal items could be the target this rules out many of the display items as distracters, shortening response time (RT) accordingly. Wolfe and colleagues found supporting empirical evidence for such visual search guidance (Wolfe, 1994a; Wolfe et al., 1989).

It is important to note that the transfer of information from the parallel to serial search process is not perfect; the visual system is not able to activate target-similar regions in separate feature maps and then sum these to reveal regions with overlapping sets of the features. Using the above example, the visual system can not activate all the red regions in the colour feature map and all the vertical regions in the orientation feature map, and sum these in the activation map such that the region with the red vertical bar has the greatest activation. If this were possible, then the serial search process would be so efficient that it would behave similarly to the parallel processing stage. Wolfe and colleagues did not find evidence to suggest that the visual system behaves in this manner. Instead, Wolfe suggested that noise led to the visual system's failure to reach full efficiency. The activation related to distracters is noise and the activation related to the target is noise plus the guiding signal: if these levels of activation are close enough then search will lose efficiency and attention can be guided away from the target object. Despite this inefficiency, the essential point to come from the *Guided Search Model* is that when the target's features are known search is more efficient than a random serial order search suggested by earlier theories.

The *Guided Search Model* has been an important contribution to the search literature. Many studies have since found evidence supporting target template guidance, using RT measures (Bravo & Farid, 2009; Egeth, Virzi, & Garbart, 1984; Kunar, Flusberg, & Wolfe, 2006; Vickery, King, & Jiang, 2005; Wolfe, 1994a, 1998; Wolfe & Horowitz, 2004; Wolfe, Horowitz, Kenner, Hyle, & Vasan, 2004), psychophysics detection studies (Beutter, Eckstein, & Stone, 2003; Burgess, 1985), eye movement studies (Castelhano, Pollatsek, & Cave, 2008; Chen & Zelinsky, 2006; Findlay, 1997; Luria & Strauss, 1975; Malcolm & Henderson, 2009, 2010; Najemnik & Geisler, 2005, 2008; Rajashekar, Bovik, & Cormack, 2006; Tavassoli, van der Linde, Bovik, & Cormack, 2009; L. G. Williams, 1967) and even in neurophysiology studies (Desimone & Duncan, 1995). The range of methodologies used in these target template studies have given researchers a better understanding of how the visual system works.

Using a Target Template to Locate Items in the Real World

A target template has been identified as a source of guidance that can help observers locate a target in a novel search display. However, research into target template guided search has reached a glass ceiling. In order to progress our understanding of how the visual system uses a target template, target template guidance needs to be understood in terms of real-world scenes. Previous empirical studies used object arrays (Castelhano et al., 2008; Chen & Zelinsky, 2006; Findlay, 1997; Hannus, van den Berg, Bekkering, Roerdink, & Cornelissen, 2006; Luria & Strauss, 1975;

Schmidt & Zelinsky, 2009; Treisman & Gelade, 1980; Treisman et al., 1977; Vickery et al., 2005; D. E. Williams & Reingold, 2001; L. G. Williams, 1967; Wolfe, 1994a; Wolfe et al., 1989; Wolfe et al., 2004; Zelinsky, 1996) or noise fields (Beutter et al., 2003; Rajashekar et al., 2006; Tavassoli et al., 2009). These displays have none of the complexities of natural environments (see Figure 4).

The paucity of research into target template guidance through real-world scene images is an important omission from the literature. As opposed to an array, scenes are "a semantically coherent (and often nameable) view of a real-world environment comprising background elements and multiple discrete objects arranged in a spatially licensed manner" (p244, Henderson & Hollingworth, 1999b). The presence of a background scene changes the information available to the observer's visual system, and thereby the observer's search strategy. A background scene allows an understanding of the gist, and thereby knowledge of the scene's probable objects and their respective locations. Conversely, a scene also introduces exponentially more features that have to be ignored when selectively guiding attention. Such conflicting effects have been empirically demonstrated to both negatively and positively affect search times (Wolfe, 1994b; Wolfe, Oliva, Horowitz, Butcher, & Bompas, 2002). Real-world scenes therefore provide their own idiosyncratic information that cannot be replicated in object arrays or noise fields. Introducing a real-world scene radically changes the nature of information available to the visual system and how that information is processed during the completion of a task.[2]

[2] An example of how substituting a scene for an array can change processing during search tasks can be found in the contextual cueing literature. Repeated exposure to a particular search array will lead to a more efficient search and significantly shorter search times than when novel arrays are displayed. This has been found to be due to recognition of local elements in the array (Jiang & Wagner, 2004), though the recognition of the array itself is implicit (Chun & Jiang, 1998). Contextual cueing can also take place in real-world scenes; however, the recognition of the display is explicit, the effect immediate, and the guidance comes from the global features (Brockmole et al., 2006; Brockmole & Henderson, 2006b). The point to be taken here is that the presence of a scene significantly alters the method with which the visual system utilises information within the display to deploy attention.

Figure 4 An example of a difference in a display when the background scene is removed, leaving only an object array. Without the scene, the semantic content of the display's objects (i.e., objects likely found in an office) and the probable spatial relationship between these objects takes longer to identify.

Researchers will not be able to understand fully the human visual search process until all sources of guidance are tested using real-world scenes. To that end, other potential sources of guidance information such as image features (Itti & Koch, 2001;

Itti et al., 1998), contextual constraints (Castelhano & Henderson, 2007; Eckstein et al., 2006; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006) and display memory (Brockmole et al., 2006; Brockmole & Henderson, 2006a, 2006b; Hollingworth, 2009) have been studied recently, while employing real-world scenes. These results have helped generate real-world search models (Ehinger et al., 2009; Itti & Koch, 2001; Itti et al., 1998; Kanan, Tong, Zhang, & Cottrell, 2009; Rao, Zelinsky, Hayhoe, & Ballard, 2002; Torralba et al., 2006; Zelinsky, 2008) which have grown increasingly more accurate in predicting human fixation positions. However, the role of a target template in guiding search in real-world scenes has not been studied. This omission from the empirical literature hinders both cognitive psychologists from fully understanding the behaviour of the visual system and computational scientists from generating accurate search models, by not providing essential data needed to constrain new search theories.

How the Addition of a Real-World Scene Might Affect Target Template Guided Search

Why can cognitive psychologists and computational modelers not just simply assume that results from previous target template studies will translate to scene search? As Wolfe demonstrated, the addition of a naturalistic background changes search performance (Wolfe, 1994b; Wolfe et al., 2002). The availability of a scene background provides additional guidance information that may make the information provided by a target template redundant. When searching through a novel object array, a target template is the only type of cognitive knowledge that can guide search. It is not surprising, then, that the availability of a target template affects RT so strongly. However, in the real world, there are other cognitive knowledge factors available, chief among them scene context. Scene context provides information regarding where the target might be located, directing eye movements (Castelhano & Henderson, 2007; Ehinger et al., 2009; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006; Torralba et al., 2006) and facilitating the verification of the fixated target object (Biederman, Mezzanotte, & Rabinowitz, 1982; Malcolm & Henderson, 2010). The visual system appears to utilise scene context information whenever it is available to guide search. This reliance on scene context information could make the information provided by a target template redundant either by providing overlapping information to the visual system as to where the target might be, or by the visual system having an information bottleneck in deciding where to fixate in a scene and simply favouring scene context to target template information. In either scenario, a target template's influence on real-world image search, as compared to object array search, would be minimal.

The other way that target template guidance might be negatively affected by the inclusion of a background scene is the increase in feature information in the display. A target template guides attention by contrasting stored features with those in the display percept. This is a straightforward process in an object array: stimuli are easy to separate from a blank background and the use of basic features makes potential targets relatively easy to distinguish from distracters. In Figure 3 example, all the elements are easy to locate, and the features used to distinguish potential targets from distracters – colour and orientation – are easily recognised. In a scene percept the blank background is replaced by a display rich in detail, containing hundreds of features across multiple dimensions at every location. These extra visual features add to the noise that the visual system has to filter, making it harder for the visual system to select appropriate regions for fixation on the basis of the target's features alone. Furthermore, the features in a scene are dynamic, changing depending on lighting and the participant's position and movement.

The ability of a target template to guide real-world visual search therefore needs to be demonstrated, and the specific search behaviours affected by a target template need to be identified, if researchers are to fully understand the real-world search process. Though a target template has been well demonstrated as a form of search guidance in stimulus arrays, its effect on real-world search guidance is not understood. Can a target template efficiently guide eye movements in the real-world given the extra complexities of a scene? And if so, which exact search behaviours are affected? Chapter 2 summarises a study investigating just this.

Chapter 2: Target Template Guidance Through Real-World Scenes

A template – the observer's mental representation of the target – can guide eye movements during search tasks (Beutter et al., 2003; Eckstein, Beutter, Pham, Shimozaki, & Stone, 2007; Findlay, 1997; Luria & Strauss, 1975; Motter & Belky, 1998; Rajashekar et al., 2006; Tavassoli, van der Linde, Bovik, & Cormack, 2007; Tavassoli et al., 2009; L. G. Williams, 1967; Zelinsky, Rao, Hayhoe, & Ballard, 1997) and facilitate perceptual decision tasks (Burgess, 1985; Greenhouse & Cohn, 1978; Judy, Kijewski, Fu, & Swensson, 1995) by increasing the weight of signals from target similar features in the display percept while de-weighting signals from target dissimilar features. However, it remains to be determined whether a target template can guide search through real-world scenes and, if so, which search behaviours are facilitated.

In order to determine this, the current study manipulated target template specificity; that is, how accurately the template represents the actual target object. The search behaviours affected by a target template should be facilitated the more specific the template becomes, while the search behaviours that are independent of the target template should not be affected by this manipulation.

Previous studies that manipulated specificity came to differing conclusions as to which search behaviours a target template affects. Wolfe et al. (2004) and Vickery et al. (2005) suggested that the target template facilitates the time needed to start search. Both studies gave participants search tasks in which the target was cued prior to the array onset (Figure 5). In order to manipulate the specificity of a target template the cue was broken up into two conditions: either a picture or word cue[3].

[3] Both studies tested more than two cue types, but for the sake of the present discussion only the exactly matching picture cue and word cue are referred to here. These two cues represent the extremes of target template guided search: cues that allow the most specific template to be generated (picture), and cues that allow the least specific template to be generated (word). For the interested reader: Wolfe et al. (2004) also used picture cues that indicated target type, but did not indicate any of the target's visual features (e.g., a picture of a German Shepherd could be cued indicating that the target was a dog, but the actual target object could be a Boston Terrier). Vickery et al. (2005)

In the picture cue condition participants were given a picture of the target exactly as it appeared in the search display. This allowed the participants to generate a specific target template in which any of its features could guide search. In the word cue condition, participants knew the identity of the object but were not given any of the visual details. As such, participants could only generate a prototypical object template, the features of which would be inefficient at guiding search.

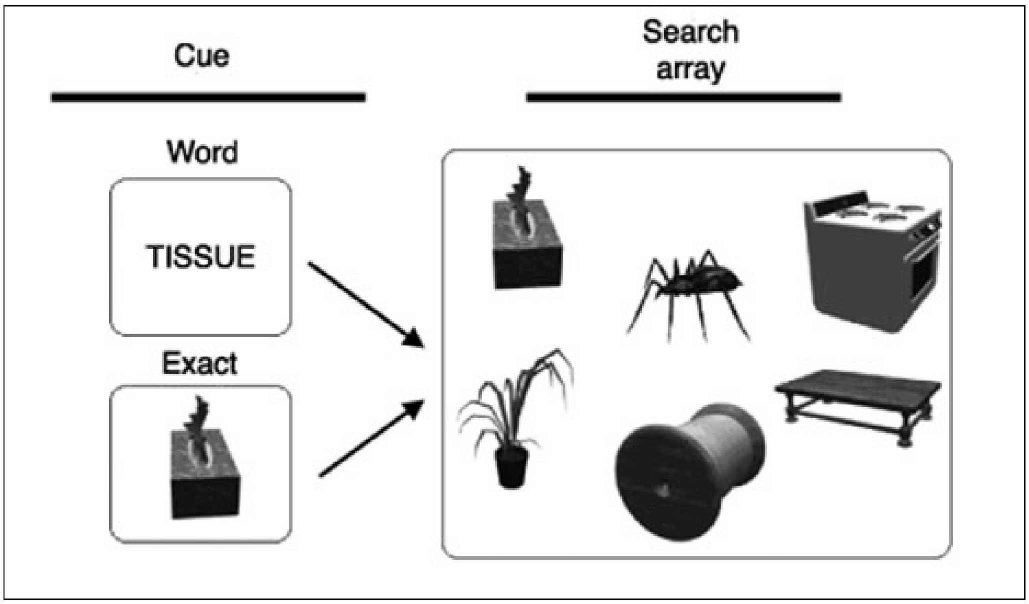

Figure 5 An example of the paradigm used in Wolfe et al. (2004) and Vickery et al. (2005). A cue identifying the target would appear, followed by a search array. (This figure was created using an image from Vickery et al., 2005; note the search array is not drawn to scale).

Unsurprisingly, RT was faster when participants were given picture cues. Surprisingly, both studies found that when the stimulus onset asynchrony (SOA) increased RT decreased in word cue trials and increased slightly in picture cued trials, leading to a significant interaction between cue type and SOA. Both studies suggested that this result was due to the time needed by the visual system to set up a target template prior to initiating search. When a picture cue was given, participants could set up a target template immediately and begin the search process. When participants were given a word cue, participants would need a longer period of time to set up a target template, after which search would begin. However, at longer

presented picture cues which manipulated target orientation (e.g., a picture cue of a motorcycle exactly as it would appear in the scene, a picture cue of a motorcycle rotated 90° in plane, a picture cue of a motorcycle rotated 90° in depth, etc.).

SOAs the word cue would have enough time to set up a target template prior to array onset, while the template generated from the picture cue would have already started to decay.

A second behaviour that a target template has been suggested to facilitate is the time to locate the target; that is, the time from the beginning of search until the target is first fixated (what this thesis will refer to as scanning time[4]). Several eye movement studies have shown that the specificity of target template information affects the visual system's selection of saccadic destinations. For example, Findlay (1997) and Rajashekar, Bovik, & Cormack (2006) found that if participants knew the shape of a target they were more likely to direct saccades to similarly shaped stimuli in the periphery. Similarly, eye movements during search are selective for orientation and spatial frequency features that matched those of the target (Tavassoli et al., 2009). Previous studies have also shown that fixations tend to land on similarly coloured distracters when participants know the target's colour (Hannus et al., 2006; Motter & Belky, 1998; D. E. Williams & Reingold, 2001; L. G. Williams, 1967). In search tasks using real-world object arrays, fewer distracters were fixated when participants were given an exact picture cue of the target than a word cue of the target, and even within word cues extra information (e.g., brown boots opposed to boots; or boots as opposed to footwear) led to a greater saccadic selectivity (Castelhano et al., 2008; Schmidt & Zelinsky, 2009). These studies collectively interpreted their results as evidence that the visual system uses target template features to facilitate the scanning epoch, specifically the saccadic selection process.

The third and final search behaviour that a target template has been suggested to facilitate is the time to verify the target object, once fixated. Burgess (1985) found that prior knowledge of a target improved identification of Hadamard signals in a static noise display during alternative forced choice tasks. Closer to the field of visual search, Castelhano et al. (2008) found that a picture cue shortened the gaze duration needed to identify the target, once located.

---

[4] Castelhano et al. (2008) analysed a similar epoch in their paper, titling this epoch target latency.

There are, however, limitations to these previous analyses that still need to be addressed. First, it is not certain that a target template facilitates the time to start search. Unlike those studies that suggest a facilitation in scanning (Findlay, 1997; Luria & Strauss, 1975; Motter & Belky, 1998; Rajashekar et al., 2006; Schmidt & Zelinsky, 2009; L. G. Williams, 1967) and verification epochs (Castelhano et al., 2008) which used an eye-tracking methodology, those studies that claimed a target template affects the search initiation epoch (Vickery et al., 2005; Wolfe et al., 2004) based their analysis solely on RT measures. This coarse measurement does not allow a fine enough analysis to prove such a hypothesis.

Secondly, the studies that claimed that saccade destination processing was facilitated with increased target template specificity did not use real-world scenes. Scenes have far more visual noise than arrays, so matching the visual features of a template with those of an image percept may not be the best form of guidance in scene search. Instead, real-world search may be dominated by scene context, a form of guidance not found in arrays. It is therefore far from certain that previous findings of target template guidance will replicate to real-world search.

Thirdly, there is another behaviour that could be facilitated during scanning which has not been addressed in previous studies. Within the scanning epoch there are multiple processes occurring at most fixations (van Diepen, Wampers, & d'Ydewalle, 1998), yet the above mentioned studies couched target template facilitation of the scanning epoch solely in terms of saccadic selectivity. This ignores the variability of fixation durations. A high percentage of fixation durations are under direct control (Henderson & Pierce, 2008; Henderson & Smith, 2009). During most fixations the visual system evaluates the fixated region and makes a decision; in the scanning epoch this decision tends to be a rejection of the fixated object as a distracter. If a target template facilitates this evaluation process, then scanning epoch fixations should be shorter.

In the current study, all these potential sites of facilitation are analysed for evidence of facilitation when a more specific template is used.

Eye Tracking Analysis Used In This Thesis

In this study, and the studies in the following chapters, an eye-tracker was used to record search behaviours. Search was divided into three behaviourally defined epochs to provide finer analysis. Each epoch reflected a separate, hypothesised underlying search process. There is precedent in eye movement research to differentiate search into epochs for finer analysis. Castelhano et al (2008) divided search into two epochs based on eye movement data: target latency or the time to search for the target, and target verification or the time to accept the target once found. In the current study, search was divided into three behaviourally independent epochs based on the eye movement record (Figure 6). First was search initiation time, defined as the time from appearance of the search scene until the first saccade away from the initial fixation point (i.e. initial saccade latency). Search initiation time is assumed to reflect the processes of establishing the search template and to select a first saccade destination. Second was scanning time, defined as the elapsed time between the first saccade (the end of the search initiation epoch) to the first fixation on the target object. This epoch was taken to represent the actual search process (cf. Castelhano et al., 2008). Third, verification time was defined as the participant's gaze duration on the target object[5], and was taken to reflect the time needed to decide that the fixated object was in fact the target. Total trial duration, the measure typically reported in visual search studies, equals the sum of these three epochs. Segmenting total trial duration into these three epochs helps elucidate the effect that target template specificity has on the search process, particularly for correct trials. Trials in which the participant fixated the target and then continued searching were removed from analyses as these cases would tend to distort the verification measures.

Within the scanning epoch, two further sub-processes were analysed. During most scanning epoch fixations there are two processes occurring sequentially (van Diepen

[5] Target objects had invisible boundaries drawn around them, extending a mean of 0.47° outside the target's edge. Any fixations within this boundary were counted as on the target object.

et al., 1998): the visual system must first process the fixated object and then, if that object is not accepted as the target, must decide where to fixate next. In order to gauge whether cue type affects either of these scanning processes, the number of scene regions visited as well as the mean scanning fixation duration was compared across conditions.

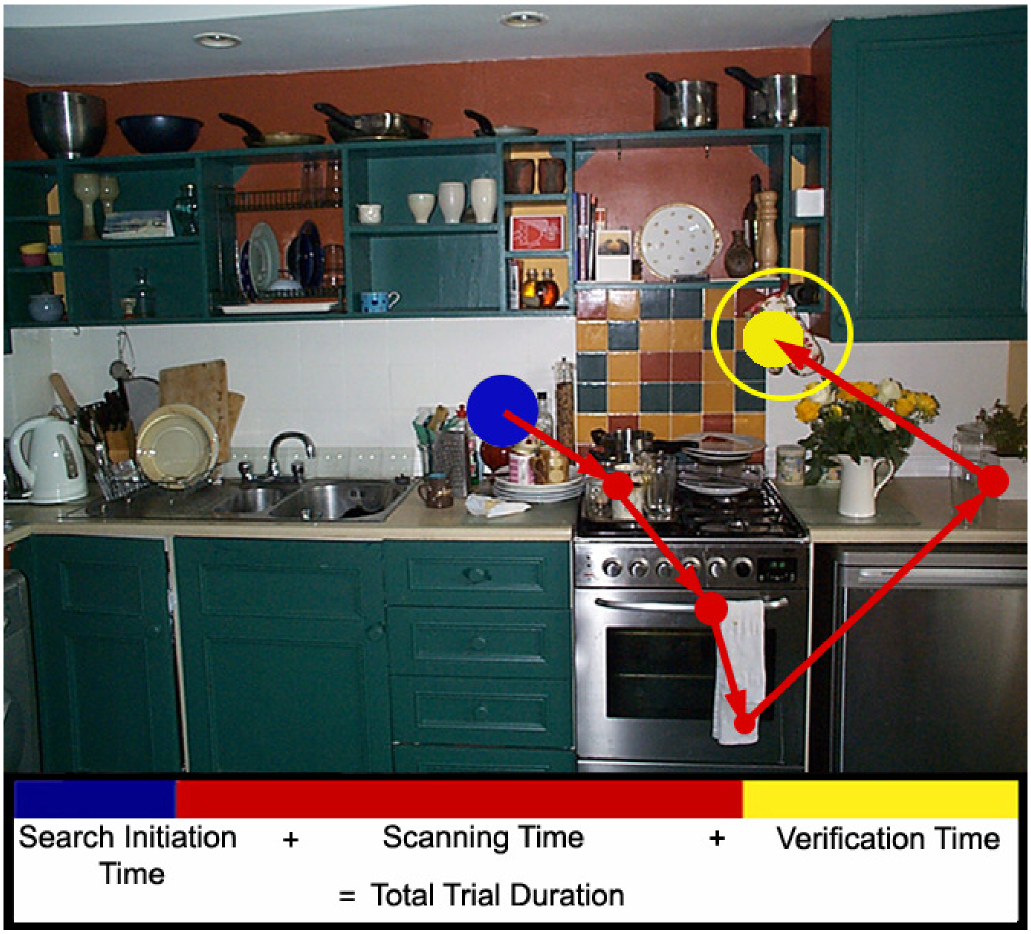

Figure 6 Dividing up visual search time. The figure displays a search scanpath. Lines represent saccades; circles represent fixations. The thin yellow line outlines the target (the oven mitt). Blue represents Initial Saccade Latency, red Scanning Time, and yellow Verification Time. When summed they yield the Total Trial Duration.

The number of scene regions visited during the scanning epoch provides an indication of whether the target cue affects the activation map. To count the number of regions visited during scanning, each scene was divided into 48 square regions of 100x100 pixels. The number of regions visited in each trial by each participant across all conditions was then calculated. Given that the measure was concerned with how many regions of a scene were visited, and not how often, regions fixated more than once were still scored as one.

The mean scanning fixation duration provides an indication of whether the target cue affects the process of matching a fixated object to an internal representation of the target. Faster processing of the item at fixation reduces the fixation duration (Henderson & Ferreira, 1990).

It should be noted that both the scanning and verification epochs contain fixations that reflect at least some similar processes such as deciding whether the target is present in the fixated region. However, there is enough of a functional difference to separate these two measures. Scanning epoch fixations involve two processes: a reject decision – which only needs a single fixated feature to mismatch with a target template – and the selection of the next fixation location. The verification epoch contains an accept decision which would likely be based on a much more complete analysis of the fixated item to ensure that it is a match to the sought target. In addition, in the verification epoch minimal emphasis need be placed on deciding where in the scene to fixate next. The difference between scanning and verification is supported by results from a recent study by Castelhano et al. (2008) which found an interaction between target typicality and cue type on verification time, but not scanning time[6].

Chapter 2's Study

In the present study participants searched for target objects in photographs of real-world scenes while their eye movements were recorded. The two manipulated properties in the experiments were the type of cue shown (word or picture) and the SOA between cue presentation and onset of the search scene (long or short).

[6] The authors found that if a target was cued with a word cue participants would locate the target just as quickly, regardless of the target's typicality (e.g., if the cue was phone, the target could be prototypical phone or a Mickey Mouse-themed phone). However, verification took significantly longer if the target had low-typicality. Target typicality had no effect during picture cued trials, either for target latency (scanning) or verification.

If a target template facilitates visual search through real-world scenes then there should be shorter total trial durations in picture cued trials. If this benefit is due to a picture cue facilitating faster target template set up in VWM prior to search, the search time interaction found in previous studies should be replicated (Vickery et al., 2005; Wolfe et al., 2004). Thus, word cues should benefit from longer SOAs, since more time is available to establish a search template prior to appearance of the scene. Picture cues, on the other hand, should suffer with longer SOAs because visually specific templates created from picture cues should decay with the extra time. However, total trial duration is a relatively coarse measure for determining target template set up time. Another more sensitive method for analysing target template set up time is provided by search initiation time. If cue specificity does affect target template set up time, search initiation times for word cue trials should be longer when the SOA is shorter. This result is expected because more time would be needed to finish establishing the template than a short SOA would allow, extending into the initial fixation once the search scene appears. Long SOAs would allow the visual system to set up the template prior to the scene's appearance. Conversely, with short SOAs, picture cues should provide a quickly available and precise template that would allow search to begin as soon as the search display appeared. Therefore, the prediction is that search initiation time should be shorter for picture than word cues with short SOAs, but this difference should be reduced or reverse as SOA increases.

If increased template specificity benefits the search process after initiation, then search time should be faster following a picture cue. Such a benefit could be revealed in scanning and/or verification epoch (cf. Castelhano et al., 2008). Within the scanning epoch cue specificity may affect the sub-processes differently. Previous models suggest that the visual system exploits the precise visual properties of a target template to improve an activation map's selection of target-probable regions (Rao et al., 2002; Zelinsky, 2008). If a target template does affect the saccade selection process then fewer scene regions should be fixated in the picture cue condition. If instead of the saccade selection process a target template affects the process of evaluating each fixated region, then mean fixation durations should be shorter in the picture cue condition.

Three experiments were conducted. Each experiment cued the search target immediately prior to the search scene, either in the form of an exactly matching picture (e.g., a picture of a coffee mug exactly as it appeared in the scene), or using a word that described the target, (e.g., the words "coffee mug"). The search scene then followed 125 to 1000ms later. Participants had to respond via a button press as soon as they found the target. Experiments 1 and 2 manipulated cue type and SOA. In Experiment 3 investigated whether target familiarity affected search by familiarising participants with all the targets prior to testing.

Experiment 1

Methods

Participants. Twelve participants gave informed consent in accordance with the institutional review board of University of Edinburgh. All participants were naïve about the purpose of the study.

Stimulus Materials. Sixty photographs of real-world scenes from a variety of categories (indoor and outdoor, natural and man-made) were used as stimuli. Search targets were chosen such that they occurred only once in the scene, did not appear at the scene centre, were not occluded, were large enough to be easily recognised yet smaller than a maximum of 3° in diameter, and were easily identifiable when presented alone (as determined by initial pilot testing).

Once the scenes were selected, they were scaled to 800x600 pixel resolution. To create the picture cues, the target objects were copied and pasted into a new blank background using Adobe Photoshop CS (Adobe, San Jose, CA). Picture cues were edited so that they did not contain any of the surrounding scene context, and then were placed at the centre of an 800x600 pixel grey background. A further 60 corresponding word cues were created that contained only the names or short descriptions of the target objects, presented in 30 point font subtending 0.89 degrees in height centred within the same grey background.

Apparatus. Eye movements were recorded using an EyeLink 1000 eye-tracker sampling at 1000 Hz. Viewing was binocular but only the right eye was tracked. Experiments were programmed in Experiment Builder. Initial data reduction was accomplished with DataViewer (SR-Research, Mississauga, ON). Stimuli were shown on a 21" ViewSonic G225f cathode ray tube monitor (ViewSonic, London, UK) positioned 90cm away from the participant, taking up an 18.72° x 24.28° field of view, with a refresh rate of 140Hz.

Procedure. Prior to the experiment, each participant underwent the EyeLink calibration procedure: Eye positions were recorded as participants fixated a series of 9 dots arranged in a square grid extending to 19.25° eccentricity. Calibration was then validated against a second set of 9 dots.

For the experiment, the trial structure was as follows. Each trial began with eye-tracking drift assessment and correction. Participants then pressed the spacebar to start the trial. A central fixation cross appeared for 400ms, followed by a cue identifying the search target for 200ms. The cue was either a word identifying the target or an exactly matching picture of the target. The cue was followed by a central fixation point lasting either 100ms or 800ms, creating two SOA conditions of 300ms and 1000ms (Figure 7). These SOA times were selected because they replicated the longest SOA duration and one of the shortest SOA durations used in previous studies (Vickery et al., 2005; Wolfe et al., 2004). The experiment was thus a 2x2 design with cue type (word vs. picture) and SOA (300ms vs. 1000ms) as the variables. Once the delay was over, the corresponding real-world scene appeared. Participants were asked to locate the target as quickly and as accurately as possible, and to press a response key as soon as the target was found. Participants were given 8 practice trials prior to the experiment.

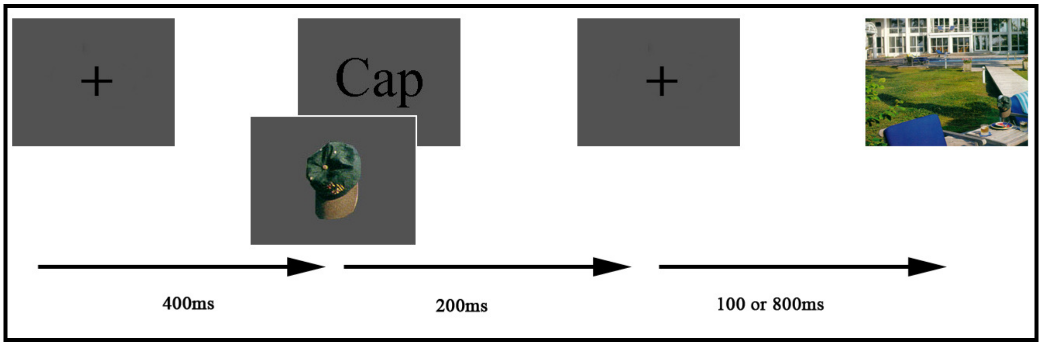

Figure 7 Procedure. A central cross is presented, followed by a cue (either a picture or word), followed by a central cross and then the search scene. Participants would then search for the target, responding with a button-press when found. SOA is measured from the onset of the cue to the onset of the search scene: either 300 or 1000ms, here. Note that the picture cue would appear the same size as the target in the scene.

Results

Trials with errors were removed from analysis; these included trials in which participants incorrectly identified the target; participants fixated the target, moved away, and returned before correctly identifying it; or the total trial duration exceeded 5500ms. Overall, 7% of trials were removed by these criteria. If a participant fixated the target once, moved fixation off the target for one fixation, and then immediately returned to the target in the next fixation, this was accepted as a correct trial. On such trials, a single fixation deviating away from the target was considered to be the result of a pre-programmed oculomotor command and not due to a decision to attend to a different possible target. This fixation sequence occurred on 4.3% of the correct trials.

Repeated-measures ANOVAs with cue type (word vs. picture) and SOA (300ms vs. 1000ms) as factors were conducted on total trial duration, search initiation time, scanning time, and verification time (Table 1).

	Word Cue, 1000ms SOA		Word Cue, 300ms SOA		Picture Cue, 100ms SOA		Picture Cue, 300ms SOA	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Total Trial Duration	1245.45	42.42	1240.65	35.46	1047.14	42.63	1146.95	80.73
Search Initiation Time	242.76	10.57	271.07	9.53	232.73	6.40	273.65	7.36
Scanning Time	500.43	34.73	483.65	35.72	374.53	30.46	436.32	65.15
Regions Visited	1.95	0.12	2.08	0.15	1.66	0.15	1.52	0.13
Fixation Durations	165.49	6.60	157.41	6.66	146.22	5.63	154.89	7.22
Verification Time	502.26	27.82	485.13	27.62	439.88	34.75	433.92	32.62

Table 1 Results from Experiment 1. All means are in millisecond units, except for Regions Visited which is measured in the number of regions visited on the display screen (maximum 48).

For total trial duration, there was a significant main effect of cue type, $F(1, 11)=19.35$, $MS_E=13215.64$, $p<0.005$, with faster response times for picture cues than word cues, indicating that search was facilitated by the ability to establish a more precise target template. SOA, however, did not produce a significant main effect, $F(1, 11)=1.57$, $MS_E=17252.93$, $p=0.236$, nor was there a significant interaction between cue type and SOA, $F(1, 11)=1.51$, $MS_E=21816.95$, $p=0.246$. Thus SOA failed to influence total trial duration.

For search initiation time, there was no main effect of cue type, $F<1$; participants began their search equivalently given a picture or a word cue. There was a significant main effect of SOA, $F(1, 11)=30.64$, $MS_E=469.24$, $p<0.001$, with a longer SOA producing quicker search initiation. Cue type and SOA did not interact, $F(1, 11)=2.240$, $MS_E=213.03$, $p=0.163$. Search initiation was therefore only affected by SOA and not the specificity of the template.

In contrast to search initiation time, cue type produced a significant main effect on both the scanning and verification epochs. Scanning and verification times were shorter for picture than word cues ($F(1, 11)=12.64$, $MS_E=7123.55$, $p<0.01$; and, $F(1, 11)=7.62$, $MS_E=5079.60$, $p<0.05$, respectively), demonstrating an advantage for a more precise target template. There was no effect of SOA in either the scanning or verification epochs (*F*s<1), and no interaction of cue type and SOA in either epoch (*F*s<1). Thus, two epochs of visual search, scanning for the target and verifying that the target had been found, were both facilitated by the more precise target cue, supporting the hypothesis that search in real-world scenes is facilitated by a more precise target template.

Since picture cues resulted in shorter scanning epochs than word cues, further analyses were conducted to specify how the cue affected the scanning processes. Specifically, the number of scene regions visited and fixation durations during the scanning epoch were examined (Table 1).

Participants visited fewer regions during scanning when they had been shown a picture cue, $F(1, 11)=14.43$, $MS_E=0.150$, $p<0.005$, consistent with the hypothesis that a more precise target template led to a more selective placement of fixations. There was no main effect of SOA, $F<1$, and no interaction between cue type and SOA, $F(1, 11)=1.475$, $MS_E=0.153$, $p=0.250$.

Scanning fixation durations were marginally shorter following picture cues than word cues, $F(1, 11)=3.778$, $MS_E=3777.348$, $p=0.078$. There was no main effect of SOA, $F<1$; nor was there an interaction between cue type and SOA, $F(1, 11)=2.292$, $MS_E=367.474$, $p=0.158$. These data suggest faster rejection of non-targets in each fixation given a more specific target template.

Discussion

The results of Experiment 1 clearly indicate that visual search in real-world scenes is facilitated by a specific target template. Total trial duration was reduced given a picture cue rather than a word cue. A more precise target template affected both scanning time and verification time, with picture cues yielding shorter scanning and verification epochs. Closer analysis of the scanning epoch revealed that picture cues allowed for fewer regions to be visited and a tendency for fixations to be shorter in duration during search. The results suggest that knowledge of a target's appearance prior to search can benefit scanning in two ways: by facilitating selection of potential target locations beyond the current fixation location, and by shortening the time needed to reject fixated distracters before moving on to the next potential target.

Interestingly, there was no interaction found between cue specificity and SOA, either in total search time or in search initiation time. Therefore, no evidence was found for either a lengthened template set up time when participants were given a word cue, or for decay of visual specificity given a picture cue. The finding that search initiation time was faster overall given a longer SOA can be accommodated by the common finding that responses are faster given more pre-trial warning-time to prepare.

Before the conclusion that there is no effect of cue specificity on search initiation time is accepted, a possible alternative explanation for the null results must be considered. It has been reported that specific cues reach close to their full advantage with SOAs around 200ms (Vickery et al., 2005; Wolfe et al., 2004). The shorter SOA in the current experiment (300ms) may have been too long to reveal an effect of cue type. Therefore, Experiment 1 was replicated but now with SOAs of 200ms and 800ms.

Experiment 2

Methods

Participants. Thirteen participants gave informed consent in accordance with the institutional review board of University of Edinburgh. All participants were naïve about the purpose of the study and none of them participated in Experiment 1.

Stimulus Materials. The stimuli were the same as Experiment 1, except the word cues' font was increased to 72 point (2.14º).

Procedure. The procedure was the same as Experiment 1 with the following exceptions. First, cues were shown for 150ms, followed by a fixation cross for 50ms or 750ms, producing SOAs of 200ms or 800ms. Second, participants now responded via a response pad (SR Research, Mississauga, ON). Third, the experimenter initiated each trial after the participant fixated a central drift-correction dot.

Results

Data from 12 participants were accepted; data from a thirteenth participant were eliminated due to an unusually high error rate (70% accuracy vs. 93% mean accuracy rate among the accepted 12 participants). Trials in which participants fixated, moved off, and returned to the target in the next fixation occurred on 3.8% of the correct trials. All correct trials were subject to four repeated-measures ANOVAs with cue type (word vs. picture) and SOA (200ms vs. 800ms) as factors, and total trial duration, search initiation, scanning time, and verification time as dependent measures. As in Experiment 1, fixation durations and number of regions visited during scanning were also analysed (Table 2).

	Word Cue, 800ms SOA		Word Cue, 200ms SOA		Picture Cue, 800ms SOA		Picture Cue, 200ms SOA	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Total Trial Duration	1246.50	49.45	1215.13	53.83	997.01	37.80	1032.19	36.95
Search Initiation Time	266.59	7.34	312.30	10.51	261.41	8.22	310.70	6.00
Scanning Time	584.95	46.06	516.29	69.10	387.63	45.70	382.18	45.43
Regions Visited	1.90	0.17	1.71	0.17	1.31	0.12	1.24	0.12
Fixation Durations	180.13	6.36	170.65	6.70	161.52	10.41	157.53	6.42
Verification Time	450.01	30.43	437.93	33.27	384.05	20.99	393.25	37.19

Table 2 Results from Experiment 2. All means are in millisecond units, except for Regions Visited which is measured in the number of regions visited on the display screen (maximum 48).

Total trial duration was shorter for picture than word cues, $F(1, 11)=33.08$, $MS_E=16957.615$, $p<0.001$, again demonstrating a search advantage when a more precise search template could be established. However, the reduced SOA still failed to produce a main effect, and it did not interact with cue type, *F*s<1.

As in Experiment 1, search initiations were faster for longer SOAs, $F(1, 11)=82.28$, $MS_E=329.105$, $p<0.001$, but there was no effect of cue type and no interaction between cue type and SOA, *F*s<1. These results are consistent with a general warning benefit from the longer SOA.

For both the scanning and verification epochs, picture cues produced shorter times than word cues, ($F(1, 11)=19.21$, $MS_E=17151.66$, $p<0.001$; and, $F(1, 11)=15.10$, $MS_E=2432.13$, $p<0.005$, for scanning and verification time, respectively). There was again no effect of SOA for either the scanning or verification epochs (*F*s<1) nor an interaction between cue type and SOA in either of the epochs ($F(1, 11)=1.21$, $MS_E=9910.41$, $p=0.295$; and, $F<1$, respectively).

As with Experiment 1, number of regions visited and fixation durations during the scanning epoch were examined to specify more precisely how the scanning epoch was influenced by the two variables (Table 2). In the regions visited analysis, fewer regions were visited following picture than word cues, $F(1, 11)=14.38$, $MS_E=0.24$, $p<0.005$. Again, there was no main effect of SOA, $F(1, 11)=1.23$, $MS_E=0.16$, $p=0.292$, and no interaction between cue type and SOA, $F<1$.

Picture cues yielded shorter mean fixation durations than word cues, $F(1, 11)=5.92$, $MS_E=510.73$, $p<0.05$, but there was no main effect of SOA and no interaction between cue type and SOA, *F*s<1.

In summary, the pattern of results remained identical to those of Experiment 1. A more specific cue facilitated search, primarily due to faster reject decisions and better targeting during the scanning epoch, along with faster acceptance of the target once it was fixated. These results demonstrate that a more precise target cue, leading to a more precise search template, facilitates search in real-world scenes.

Discussion

The results of Experiment 2 were almost identical to those of Experiment 1, with total trial duration, scanning time, and verification time all facilitated by a specific picture cue. The shorter scanning epoch again resulted from significantly faster reject decisions at each fixation and a better targeting of the next fixation. Search initiation was faster following longer SOAs. However, there was no interaction between cue type and SOA on either total trial duration or search initiation time.

Prior studies reported an interaction of cue type and SOA on RT, hypothesized to be partially a result of more time needed to establish a target template following a word cue than a picture cue (Vickery et al., 2005; Wolfe et al., 2004). Why was this interaction not replicated, either in the total trial duration or search initiation measure? One possible explanation is that search initiation time is independent of cue specificity and that a longer SOA may simply provide more general warning that the trial will begin. However, this does not explain the lack of an interaction of factors on total trial duration. Another possible explanation for the null effect in the current results, not addressed in the first two experiments, is related to participants' overall familiarity with the targets. In Vickery et al. (2005), participants were familiarised with the appearance and name of all targets prior to testing. Similarly, in both that study and the study by Wolfe et al. (2004), each target was presented several times during the experiment. It is possible that in those studies, participants used word cues to retrieve from visual long term memory (VLTM) an image of the target learned during the experiment, either before initiating search or during the course of search. This might be a time-consuming process that could be completed prior to the trial given a longer SOA, but that might be less likely to be completed until after the trial had started given a shorter SOA. Picture cues, in contrast, would not require any retrieval from VLTM, and could be used to establish a specific template in the duration allowed by a short SOA.

Experiment 3 investigated this possibility by familiarizing participants with targets prior to the experiment. If the failure to find evidence for varying target template set up time in Experiments 1 and 2 was due to lack of familiarity with the pictorial properties of the targets in the word cue condition, then an interaction between cue specificity and SOA in Experiment 3. This would be demonstrated in either the total trial duration or search initiation time measures.

It is also possible that 200ms is still not a short enough SOA to reveal an interaction of SOA and cue type. Previous reports suggest that most of the advantage of real-world picture cues can be gained by an SOA of around 200ms (Vickery et al., 2005; Wolfe et al., 2004), so Experiment 3 decreased the shorter SOA to 125ms: a duration short enough that it should probe a period of target template set up and long enough that the word cues are still identifiable.

Finally, one might argue that search in the first two experiments was too easy, leading to a ceiling effect that might mask an effect of SOA on scanning. Participants found the targets relatively quickly with total trial durations means in the range of 1000-1250ms. Participants in previous research took up to 1700ms (Vickery et al., 2005). Therefore all scenes with mean total trial durations less than 950ms in Experiments 1 and 2 were replaced with more difficult search scenes in Experiment 3.

Experiment 3

Methods

Participants. Fifteen participants gave informed consent in accordance with the institutional review board of University of Edinburgh. All participants were naïve about the purpose of the study and none of them participated in Experiments 1 or 2.

Procedure. Experiment 3 followed the same procedure as Experiments 1 and 2 with the following exceptions. First, participants were shown all possible target pictures together with their associated words four times each prior to the experiment. Target picture-word pairs were shown in random order, and were self-paced. Participants were told to pay attention to this learning session because target cues during the experiment would be presented briefly and it would benefit them to know a target's appearance in advance. Second, the target cue was displayed for 75ms followed by a fixation cross for 50ms in the short SOA condition and 725ms in the long SOA condition. This resulted in SOAs of 125 and 800ms.

Stimulus Materials. All scene images from Experiment 2 with mean total trial duration across participants below 950ms or that contained human faces were replaced in Experiment 3. This resulted in replacement of 24 scene images and their corresponding picture and name templates. Selection of new scenes followed the same criteria as before.

Results

Three participants were removed from the analysis, one due to poor calibration and two for failing to follow instructions. Data from 12 participants were analysed. There was a 92% mean accuracy rate among these 12 participants. Trials where participants fixated, moved off, and returned to the target within one fixation occurred on 3.5% of the correct trials. Data analyses mirrored Experiments 1 and 2 (Table 3).

As in Experiments 1 and 2, total trial duration was faster following picture cues, $F(1,11)=40.78$, $MS_E=22593.65$, $p<0.001$. In addition, total trial duration was also shorter following long SOAs, $F(1,11)=5.46$, $MS_E=19585.81$, $p<0.05$. However, there was still no interaction between the two variables, $F<1$.

Search initiation time was faster following longer SOAs, $F(1,11)=44.72$, $MS_E=1136.03$, $p<0.001$, just as in Experiments 1 and 2, but again there was no significant main effect of cue type, $F<1$, and no significant interaction between cue type and SOA, $F(1,11)=1.41$, $MS_E=680.54$, p=0.260. Thus, even when participants were familiarised with target appearances prior to the experiment, SOA only affected general rather than cue-specific preparatory processes.

	Word Cue, 800ms SOA		Word Cue, 125ms SOA		Picture Cue, 800ms SOA		Picture Cue, 125ms SOA	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Total Trial Duration	1411.76	36.13	1507.12	74.19	1135.62	56.77	1229.04	58.59
Search Initiation Time	264.52	17.11	320.64	11.03	254.92	11.81	328.94	16.65
Scanning Time	772.15	38.13	838.08	63.49	580.24	51.83	569.44	36.68
Regions Visited	2.59	0.20	2.41	0.17	2.15	0.19	2.32	0.15
Fixation Durations	179.78	5.17	182.02	7.14	165.51	6.61	171.31	6.72
Verification Time	375.09	17.71	348.39	22.57	300.45	18.04	330.65	19.32

Table 3 Results from Experiment 3. All means are in millisecond units, except for Regions Visited which is measured in the number of regions visited on the display screen (maximum 48).

Both scanning and verification time epochs were shorter for picture- than word-cued trials ($F(1,11)=44.15$, $MS_E=14410.98$, p<0.001; and, $F(1,11)=33.41$, $MS_E=766.19$, p<0.001, respectively). Even with the shorter SOA of 125ms, there was still no effect of SOA on either the scanning or verification epochs (Fs<1). There was also no interaction between cue type and SOA in the scanning epoch ($F<1$), and only a marginal trend towards an effect in the verification epoch, $F(1,11)=3.29$, $MS_E=2948.79$, p=0.097.

The number of regions visited and duration of fixations during the scanning epoch were examined as a function of cue type and SOA (Table 3). Fewer regions of a scene were visited following picture cues, $F(1, 11)=5.64$, $MS_E=0.15$, $p<0.05$. There was no main effect of SOA, $F<1$, nor a significant interaction between cue type and SOA, $F(1, 11)=2.24$, $MS_E=0.16$, $p=0.163$. Fixation durations were shorter for picture than word cued trials, $F(1, 11)=30.54$, $MS_E=139.16$, $p<0.001$, and for longer than shorter SOAs $F(1, 11)=5.37$, $MS_E=75.37$, $p<0.05$. However, cue type and

SOA did not interact, $F<1$. These results replicate the results from Experiments 1 and 2.

Discussion

By shortening the short SOA condition to 125ms, the SOA variable now produced an effect on total trial duration. However, despite this effect and despite the fact that participants were familiarised with the search targets prior to the experiment, there was still no interaction between cue type and SOA on either total trial duration or on search initiation time. These results appear most consistent with the hypothesis that search initiation time reflects a general preparatory process rather than the time needed to set up a target template.

At the same time, cue specificity affected total trial duration, as in Experiments 1 and 2. This effect was seen in the scanning and verification epochs, with picture cues producing shorter times. Within the scanning epoch, picture cues were again found to reduce the number of regions visited and shorten the fixation durations. Even though participants were familiarised with the visual forms of the search targets prior to the experiment, a more specific pictorial cue presented just before scene onset facilitated search. The facilitation was primarily due to faster reject decisions and better targeting during the scanning epoch, and faster target verification after it was fixated.

Again the results failed to replicate the cue type x SOA interaction found in Vickery et al. (2005) and Wolfe et al. (2004). Since participants were familiarised with targets prior to the experiment rather than repeating targets several times over the course of the experiment, it is possible that participants did not remember targets as well as participants in the above studies. However, all participants reported post-hoc that they clearly remembered target appearances, even after one repetition of the familiarisation phase. Secondly, these results are consistent with visual search studies reporting that participants prefer searching for a target rather than recalling it (Oliva, Wolfe, & Arsenio, 2004; Wolfe et al., 2004).

General Discussion

Humans constantly search for task relevant objects to assist them in their daily lives. Research on gaze control during novel scene viewing has tended to focus on two sources of information: image features operating in a bottom-up manner (Itti & Koch, 2000; Itti et al., 1998; Parkhurst et al., 2002) and cognitive knowledge structures, particularly scene context, operating in a top-down manner (Castelhano & Henderson, 2007; Eckstein et al., 2006; Neider & Zelinsky, 2006).

Chapter 2 investigated how a target template facilitated real-world search by analysing three independent, behaviourally-defined epochs of search, as well as their respective eye movement behaviours.

Three experiments confirmed that picture cues reduce total trial duration, replicating with real-world scenes previous studies that used object arrays (Castelhano et al., 2008; Schmidt & Zelinsky, 2009; Vickery et al., 2005; Wolfe et al., 2004). Furthermore, by using eye movement measures to divide the search process into functional epochs, it was found that the shorter total trial duration in the picture cue condition was due to facilitated scanning and verification times (see also Castelhano et al., 2008, for similar results in object arrays), but not the time needed to set up a target template prior to search. Within the scanning epoch, a specific target template reduced the number of scene regions the participant visited and also the mean fixation duration.

It has been known for some time that target features can guide visual search in a top-down manner (Wolfe, 1994a; Wolfe et al., 1989). For example, Wolfe's *Guided Search Model* posited that a stored representation of the search target modulates low-level feature maps to enhance particular channels; for example, if the target is red, a target template can be used to enhance the activation of all red items in the colour feature map. When the feature maps are summed they form an activation map that highlights regions in the visual display with target-similar features. The more definite the target template the more selective the feature maps, and thereby the activation map, can be. However, the *Guided Search Model* does not explain how attention is sequentially distributed during a search task, an essential concern when considering eye movements in information-rich real-world scenes.

Recent real-world search models predict that the visual system exploits precise visual properties of a target template to improve selection of peripheral locations on the activation map, with highly activated regions drawing attention and eye movements (Kanan et al., 2009; Rao et al., 2002; Zelinsky, 2008). The present data extend these models in two particular ways. First, they show that target template specificity affects how well eye movements are distributed during search. The above models suggest that an exact target template affects attention distribution, but without manipulating the specificity of the template these studies can not pinpoint the exact behaviours facilitated by the template. Secondly, these models predict that a target template will benefit search by modulating the activation map to improve the selection of fixation locations, but make no predictions about the time required for processing potential targets once fixated (but see Zelinsky, 2008). The present study indicates that the processing of fixated objects, whether they are the target or a distracter, is faster when a specific template is available for comparison.

The current results also extend the results from the Castelhano et al. (2008) study. That study similarly demonstrated that a picture cue facilitates the scanning epoch, but did not address why this would be apart from participants needing fewer fixations to reach the target. Fixation count, however, is an ambiguous measure: fewer fixations could mean that an exact target template improves selectivity in the activation map so that more probable targets are fixated and less probable ones are ignored, thus reducing scanning time. Alternatively, fewer fixations could mean that an exact template allows attended distracters to be processed faster once fixated. Faster processing at fixation reduces the chances of a re-fixation occurring (Henderson & Ferreira, 1990), and fewer fixations would reduce scanning time. The current study finds evidence for both possibilities: with an exact template participants visited fewer scene regions during scanning, indicating a better ability to select probable target regions; and scanning fixations were shorter, indicating that distracters were matched with the template and rejected quicker.

Another key feature of the results was the failure to replicate the interaction between cue type and SOA, found in both Vickery et al. (2005) and Wolfe et al. (2004). In these studies, trials cued with a word led to longer RTs than trials cued with a picture. Yet at longer SOAs, RTs for trials cued with a word were significantly reduced, while those cued with pictures grew slightly longer. The interpretation was that a target template created from an abstract cue takes longer to set up before becoming fully useful, whereas a precise target template created from a picture cue can be established more quickly but decays over time. Each of the previous studies, and the current one, used a similar type of categorical cueing in the word cue condition (e.g. the word cue apple as opposed to fruit), suggesting that the failure to replicate the previously found interaction was unlikely to be due to different text labels. An obvious difference between the studies of Vickery et al. (2005) and Wolfe et al. (2004) and the current one is that both previous studies used objects arrays – sets of objects with no established spatial relations between each other – whereas the present study used semantically rich real-world scenes. In object arrays the only top-down information available to guide search is knowledge of the target's appearance, so RTs might be sensitive to any changes to the specificity of the target template. In real-world scenes other information is available to guide search (e.g. scene context: Castelhano & Henderson, 2007; Eckstein et al., 2006; Malcolm & Henderson, 2010; Neider & Zelinsky, 2006). A potential reason for not finding an interaction between cue type and SOA in the present study is that when a participant knows that other information will be available in the scene to guide search, they may devote fewer cognitive resources to creating a detailed target template. Participants may simply store a coarser version of the template – either extracting and storing fewer features from a picture cue or forming a less specific template from a word cue – leaving the template less susceptible to changes in SOA.

A second difference between the current study and those of Vickery et al. (2005) and Wolfe et al. (2004) is that the current study used an eye tracker while the others only required participants to make a button-press. The fact that participants were aware that they were having their eye movements recorded could have abnormally affected search behaviours. However, to the author's knowledge there is no study that documents a change in eye movement behaviours when participants are aware that their eye movements are being recorded versus when they are not; moreover, given the speed of RTs in the current study, it is unlikely eye movements were conscientiously affected in this manner.

The present results also appear to contrast with those of Foulsham & Underwood (2007), who found no difference in search time or fixation count between categorical- and instance-cued search. The two studies differed, however, in that Foulsham & Underwood (2007) used a single word cue prior to a block of trials, either indicating a categorical target type (in their case fruit, meaning that several different types of fruit could be the target over the block of trials) or an instance cue (for example, apple, meaning that the target for every trial in the block would be an apple). Participants were never given an exact picture of a cue, and since the angle, size, luminance, colour and other features of the target naturally changed from trial to trial, it would be impossible for their participants to generate a specific target template. Even if participants could store a few features from trial to trial in the instance cue block (e.g. it was a red apple), the current results (Experiment 3) and those of previous experiments (Wolfe et al., 2004) indicate that seeing a picture cue of the target is more effective than recalling it.

Isolating the Subset of Features Necessary for Guidance: The Dominance of Colour Information

The results in Chapter 2 indicate that target template specificity affects real-world search behaviour. This was found by cueing participants with the two extremes of specificity: a picture cue exactly as the target appeared in the display, and a word cue that only revealed the identity of the object. While this study reveals how template specificity affects search performance, it does not reveal which features the visual system is using to guide search. A visual system could utilise every visual feature available in an exact target template to guide search, but a more efficient method would be to utilise only a subset of features. It is impossible from the current study, however, to identify which features might be preferred by the visual system in guiding real-world search since all the features were present in the picture cue condition and none in the word cue condition.

Previous studies have identified individual features that the visual system can rely on to guide search: colour, shape, orientation, spatial frequency, etc. However, it has been suggested that colour, in particular, is preferred by the visual system over other features when guiding search. Several studies have indicated that saccade selectivity is affected more by colour information then that of orientation, size and shape (Hannus et al., 2006; Keech & Resca, 2010; Luria & Strauss, 1975; Motter & Belky, 1998; D. E. Williams & Reingold, 2001; L. G. Williams, 1967).

Chapter 3 investigates whether the visual system relies primarily on colour information to guide search through real-world scenes and, if so, identifies which particular search behaviours are most effected. In order to investigate this Chapter 3 uses the same picture and word cue conditions as in Chapter 2, but also includes a picture cue with colour information manipulated. The respective results are then compared with those coming from the other two conditions. If colour is a feature preferred by the visual system when guiding search as previous studies suggest, then manipulating colour information should have a large effect on search time and to the particular behaviours it facilitates, while manipulating the availability of other features (size, orientation, etc.) should have a negligible effect.

Combining Target Template With Other Guidance Information Types to Facilitate Real-World Search

Chapter 2's results indicate that a target template can be added to the list of information types that can guide search through novel, real-world scenes. However, most research to date has focused on how the visual system processes these forms of information individually. For instance, the processing of image properties has tended to be studied in isolation in the saliency model (Itti & Koch, 2000; Itti et al., 1998). Similarly, the effect of scene context has tended to be studied in isolation (Castelhano & Henderson, 2007; Eckstein et al., 2006; Neider & Zelinsky, 2006), and in Chapter 2 the effect of target template specificity was studied in isolation. In the real-world, however, when all these forms of information are available, a more efficient method of guiding visual attention would be to integrate two or more information sources during a search task.

The benefit of integrating different processes in real-world search has been demonstrated in the *Contextual Guidance Model* (Torralba et al., 2006). In this model saliency at a local spatial scale is constrained by scene context at a global spatial scale. Areas of high salience within a selected global region are given higher weights on an activation map than those that fall outside the selected global region. The model accurately predicted participant's first few eye movements in a counting search task. This suggests that the visual system benefits from integrating multiple sources of information.

Recent empirical evidence, however, indicates that the visual system relies less on low-level saliency than the contextual guidance model suggests (Einhäuser et al., 2008; Foulsham & Underwood, 2007; Henderson et al., 2007; Henderson et al., 2009; Zelinsky et al., 2006). A future integrated real-world search model may benefit from substituting a form of top-down information, such as a target template, in place of saliency at the local spatial scale. Instead of selecting where within a selected global scene region to fixate based on saliency, a model could compare the target template stored in VWM with the incoming scene percept to weight the activation map accordingly. Local regions with high correlations to the target template are given higher weights on the activation map, particularly if they fall within the selected global region. This way the visual system is looking for target-similar features within a target-probable region of the scene, rather than highly salient features which may not have any correlation with the target object.

Using a similar paradigm and set of analyses as Chapter 2, Chapter 4 investigates if the visual system actively combines target template with scene context information to guide real-world search.

Conclusion

The results from Chapter 2 indicate that a target template can guide attention during real-world visual search. A more specific target template facilitates scanning for and verification of the target. Search initiation time appears to be an automatic process affected only by SOA, with longer SOAs producing shorter initial saccade latencies. Familiarizing participants with the target prior to testing did not affect the target template guidance process.

The current study also demonstrates that eye-tracking allows insight into the processes underlying real-world visual search. By separating traditional unitary RT measures into three behaviourally defined epochs the underlying processes that are affected by target template specificity were examined, increasing our knowledge of search processes.

Chapter 3: Colour Information's Effect on Target Template Guided Real-World Search

There are several different features in a target template which the visual system can use to guide search: orientation, size, shape, spatial frequency, etc. This range attests to the visual system's flexibility in utilising available information to guide search.

However, if the observer has a template stored in visual working memory (VWM) with multiple visual features available, which features will the visual system rely upon for different search behaviours? The visual system could guide search by weighting every available feature equally. This would mean that every available visual feature is used to select saccade destinations, to evaluate fixated regions, and so forth. However, given the number of potential features available and the number instances in search that can be facilitated, using every feature equally would become computationally demanding and inefficient. A more efficient method would be for the visual system to rely primarily on certain features for specific search behaviours. An optimal system might, for instance, make saccade destination decisions based on colour and size of the target template, but might rely more on orientation and shape for the verification of the fixated object.

There is some supporting evidence for this hypothesis, particularly with regards to the process of selecting saccade destinations. In particular, colour information has been identified as a feature that the visual system prefers to use when weighting the activation map. Empirical studies have shown that cueing a target's colour affects the saccadic decision process more than size, shape and orientation. Even when size, shape or orientation information is added to an already coloured-defined target template, saccadic selection improves only marginally with the majority of fixations still landing on similarly coloured objects (Hannus et al., 2006; Motter & Belky, 1998; D. E. Williams & Reingold, 2001; L. G. Williams, 1967; but see Findlay, 1997)[7].

Since previous studies indicate that colour is preferred over other features when guiding search through arrays, and this thesis is concerned with understanding target template guidance in real-world scenes, it is important to know a) whether template colour information is also preferred by the visual system when looking through scenes and b) which particular search behaviours – of those identified as being facilitated by a target template in Chapter 2 – are affected by colour information.

In relation to the first query, previous studies had participants search through arrays of simple shape stimuli, with the exception of Schmidt and Zelinsky (2009). The danger – similar to the caveats raised in Chapter 2 – is that the influence of colour information may have been over represented in these previous studies. As opposed to searching through an array of stimuli on a blank background where easily detectable colours clearly distinguish potential targets from obvious distracters, scene images consist of a spectrum of colours that differ at almost every location due to different physical elements and inconsistencies of lighting. This background noise may mean that colour – though still a feature that can guide search – is not so dominant that other target features are rendered practically irrelevant during a search task. Instead, the visual system may have to rely more on other features, as well as other cognitive factors (e.g., scene context) meaning that colour is no longer a dominant source of guidance (see D. E. Williams & Reingold, 2001).

Secondly, if colour is indeed a dominant feature in guiding search through real-world scenes, it is important to identify which specific behaviours are being facilitated. Though target templates have been suggested to facilitate the time to start search, the time to locate a target, and the time to verify it, all previous empirical (Hannus et al., 2006; Motter & Belky, 1998; D. E. Williams & Reingold, 2001; L. G. Williams, 1967) and modeling studies (Itti & Koch, 2000; Navalpakkam

[7] Findlay (1997) found that participants were more likely to fixate target-similar distracters if they shared a similar shape rather than colour. However, these experiments only looked at the landing position of the first saccade and not of the entire search process.

& Itti, 2005; Rao et al., 2002; Zelinsky, 2008) discuss colour's role in search in terms of saccade selection. There are several other search behaviours that colour information could facilitate that remain unidentified.

Chapter 3's Study

In order to investigate the role of colour information during search, participants were given real-world search tasks while their eye movements were recorded. Search time was then segmented into search initiation time, scanning time, and verification time. As in Chapter 2, participants were cued with a picture or word. These cues allowed participants to generate target templates of highest and lowest specificity, respectively (see Chapter 2; Castelhano et al., 2008; Malcolm & Henderson, 2009, 2010; Vickery et al., 2005; Wolfe et al., 2004). A third condition was added in which colour was the only feature removed from an otherwise complete picture cue, creating a grey-scaled target template condition (Figure 8).

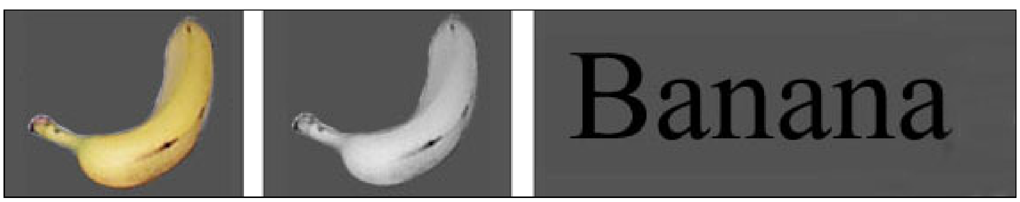

Figure 8 Example of three possible types of cue shown in Experiment 4. The colour picture and the word cues were the same as the cues used in Chapter 2's experiments. The middle picture represents the additional cue type used here: a grey-scale picture cue.

Word and picture cues were used in the present study to demarcate the extremes of template specificity's effect on search, and judge the relative influence of colour by seeing where grey-scale cued trials fall between the two. This format of using word and picture cue conditions as relative markers and having a third condition that manipulates one feature to see that feature's relative importance to target template guidance has been used before to great effect by Vickery et al. (2005) when studying manipulations to a target template's orientation.

If colour is as dominant a guidance feature as previous studies suggest, then removing colour information from a target template should have a negative effect, with the grey-scale condition search times becoming significantly slower than those from the picture condition, and similar to those generated of the word cue condition. Alternatively, if a target template's colour information is of minimal importance when searching through scenes, its removal should have minimal effect on the search process: the resulting search times from a grey-scale cue would remain similar to those of the coloured picture cue condition and remain significantly faster than a word cue condition. If a target template's colour information plays a role in guiding real-world search, but is one among many features used, removing it should affect the search process a small but significant amount. In this case search times in grey-scale picture cued trials should end up being significantly longer than colour picture cued trials, but still significantly faster than those of word cue trials.

The other goal of Chapter 3 is to identify which specific search behaviours were facilitated by a target template's colour. The previous empirical (Hannus et al., 2006; Motter & Belky, 1998; D. E. Williams & Reingold, 2001; L. G. Williams, 1967) and modeling studies (Itti & Koch, 2000; Navalpakkam & Itti, 2005; Rao et al., 2002; Zelinsky, 2008) all discuss colour's role in search in terms of saccade selection: that is a scene region is selected for fixation based on its correlation with the colour of the target object. However, this only makes up one aspect of search. Search is a spatiotemporal process that unfolds over time. As well as seeing if the visual system uses colour information to select saccade destinations in real-world images, the current study investigates whether colour information improves other search processes: in particular the process of evaluating the fixated region, both when making a reject (when fixating distracters) and accept decisions (when fixating the target). Measuring both scanning fixation durations and the verification time indicates whether the target template affects the process of matching a fixated object to an internal representation of the target. Faster processing of the item at fixation would reduce the time needed.

If colour is a preferred feature used by the visual system when directing gaze, as the above studies have indicate in stimulus arrays, then there should be fewer scene regions visited in the colour picture than the grey-scale cue condition. If colour is a preferred feature used by the visual system in evaluating the fixated object, then there should be shorter scanning fixations, and a shorter verification epoch in the colour picture than the grey-scale cue condition.

Experiment 4

Methods

Participants. 18 participants were run; all gave informed consent in accordance with the institutional review boards of University of Edinburgh.

Stimulus Material. Sixty photographs of real-world scenes from a variety of categories (indoor and outdoor, natural and man-made) were used as stimuli. One object was chosen as the search target from every scene. Criteria for target selection included: that the target was coloured (i.e., not white, black or grey), the target type only occurred once in the scene, did not appear in the centre of the image, was not occluded, and was large enough to be easily recognised yet smaller than a maximum of 3° in diameter. Each scene was scaled to 800x600 pixel resolution.

Picture cues were created by copying target objects and pasting them into a new blank background using Adobe Photoshop CS (Adobe, San Jose, CA). Picture cues were edited so that they did not contain any of the surrounding image, and were then positioned at the centre of an otherwise empty 800x600 pixel image. Using Adobe Photoshop, colour picture cues were converted to *Lab* colour space which represents colour independent from luminance along the *a* dimension, extending from green to red, and the *b* dimension, extending from blue to yellow. By using *Lab* colour space colour information in both the a and the b channels could be deleted without affecting physical luminance levels (dimension *L*). A further 60 corresponding word cues were created that contained only the names of the target objects, presented in 72 point font subtending 2.14 degrees in height centred within the same grey background. All colour picture, grey-scale picture and word cues were immediately recognisable according to initial pilot testing and post-hoc reporting from Experiment 4's participants.

Apparatus. Eye movements were recorded using the same eye-tracker as in Chapter 2.

Procedure. A central fixation cross was shown for 400ms, followed by either a colour picture, grey-scale picture or word cue for 800ms, followed by a central fixation cross for 200ms and then a full-coloured, real-world scene image. There were 60 trials, 20 for each condition, shown in a random order. Each participant was given 9 practice trials prior to testing, containing 3 trials from each condition.

Results

All 18 participants' data were included in the analysis. Two trials were removed across all participants for having accuracy rates less than 66%. Accuracy ranged across participants from 84.4-100% with a mean accuracy score of 93.4%. All incorrect trials were removed from analysis. Incorrect trials included situations where participants incorrectly identified the target, the trial exceeded 6000ms, or when participants fixated the target, moved away, and returned before correctly identifying it. Trials where participants landed on the target, moved off the target and then immediately returned with the next fixation were accepted as correct. On such trials, a single fixation deviating away from the target was considered to be the result of a pre-programmed oculomotor command and not due to a decision to attend to a different possible target. This occurred on 4.8% of correct trials.

The results showed a main effect of cue type on total trial duration, $F(2, 34)=32.078$, $MS_E=14156.925$, $p<0.001$ (

Table 4). A planned contrast found a strong trend toward a significant difference between occasions when a word and grey-scale picture cue was used, $F(1, 17)=4.009$, $MS_E=22820.586$, $p=0.061$, with grey-scale picture cues leading to shorter total trial durations. There was also a significant difference between the coloured picture and grey-scale cue condition with coloured picture cues having shorter total trial duration: $F(1, 17)=38.831$, $MS_E=25645.113$, $p<0.001$.

	Word Cue		Grey-Picture Cue		Colour-Picture Cue	
	Mean	**SE**	**Mean**	**SE**	**Mean**	**SE**
Total Trial Duration	1379.26	54.98	1307.96	56.93	1072.75	59.25
Search Initiation Time	242.03	8.39	240.53	10.19	225.38	9.50
Scanning Time	645.06	37.88	593.69	28.93	441.31	37.30
Regions Visited	2.54	0.12	2.38	0.12	2.02	0.11
Fixation Durations	174.06	6.74	177.63	7.36	161.14	7.62
Verification Time	492.22	38.06	474.50	42.02	405.95	33.86

Table 4 Results from Experiment 4. All means are in milliseconds except for Regions Visited which is measured in the number of regions visited on the display screen (maximum of 48).

The three independent search epochs that add up to the total trial duration were then analysed. There was a main effect of cue type on search initiation time, $F(2, 34)=4.750$, $MS_E=312.500$, $p=0.015$. A planned contrast revealed no significant difference between occasions when a word and grey-scale picture cue was used, $F<1$. However, the colour picture cue condition had significantly shorter latencies than the grey-scale picture cue condition: $F(1, 17)=8.276$, $MS_E=709.791$, $p=0.010$.

There was a main effect of cue type on scanning time, $F(2, 34)=13.229$, $MS_E=15277.941$, $p<0.001$. A planned contrast indicated no significant difference in scanning time between the word cue and the grey-scale picture cue condition, $F(1, 17)=1.928$, $MS_E=24646.147$, $p=0.183$. However, scanning time in the colour picture cue condition was significantly shorter than the grey-scaled picture cue condition: $F(1, 17)=15.814$, $MS_E=26427.593$, $p=0.001$.

There was also a main effect on verification time, $F(2, 34)=13.676$, $MS_E=2732.026$, $p<0.001$. Similar to the scanning epoch, planned contrasts in the verification epoch indicated that there was no difference in the time to verify the object in the grey-scale picture and word cue conditions, $F<1$, but that verification time was significantly shorter in the colour than the grey-scale picture cue condition, $F(1, 17)=12.772$, $MS_E=6621.765$, p=0.002.

The Underlying Behaviours Affecting Scanning Time

Since there was a significant difference between grey-scale and colour picture cue conditions, further analyses were conducted to see if the differences in the scanning epoch were caused by the same eye movement behaviours.

There was a main effect of cue type on the number of regions visited during scanning time, $F(2, 34)=8.350$, $MS_E=0.151$, $p=0.001$. A planned contrast revealed no significant difference between occasions when a word and grey-scale picture cue was used, $F(1, 17)=1.365$, $MS_E= 0.338$, p=0.259. However, there was significantly fewer regions visited in the colour picture condition than the grey-scale condition, $F(1, 17)=8.498$, $MS_E= 0.270$, p=0.010.

There was a main effect of cue on scanning epoch fixation durations, $F(2, 34)=7.462$, $MS_E=181.493$, $p=0.002$. A planned contrast revealed no significant difference between fixations in the grey-scale picture and word cue conditions, $F(1, 17)=1.1248$, $MS_E= 183.783$, p=0.279. However, there was a significant difference between the grey-scale and coloured picture cue conditions, $F(1, 17)=13.411$, $MS_E=$ 364.838, p=0.002, with colour picture cue trials having shorter fixations.

Discussion

Experiment 4's results indicate that removing colour information from an otherwise specific picture cue drastically reduced search efficiency, so much so that search time

was almost as slow as search with a word cue (
Table 4). This is a huge reduction. Previous studies consistently found that participants prefer a visual depiction of the target over a semantic cue (Castelhano et al., 2008; Malcolm & Henderson, 2009, 2010; Vickery et al., 2005; Wolfe et al., 2004). The current study shows that when colour information is removed from a pictorial cue, search becomes as nearly inefficient as when using a semantic cue. This supports previous studies in indicating the visual system's preference for colour information in guiding search (Hannus et al., 2006; Luria & Strauss, 1975; Motter & Belky, 1998; D. E. Williams & Reingold, 2001; L. G. Williams, 1967) and extends this documented dominance from search through stimulus arrays to real-world scene images.

Colour's facilitation to the guidance process appears to be a result of facilitation to both the scanning and verification epochs. During scanning, removing colour information from an otherwise complete picture cue increased scanning time so much that it was no longer significantly faster than search guided by a semantic cue. This increase was found to be due to an increase in the number of regions visited and the mean fixation duration. These results suggest that the colour channel plays a dominant role in weighting the activation map, supporting previous research, and is also one of the primary features used in evaluating fixated regions.

Similar results were also obtained in the verification epoch where colour picture cues allowed for fastest verification time, but removing colour information from a picture cue slowed down the rate of verification to the level provided by a semantic cue.

Search initiation was also faster for a colour picture cue than a grey-scaled picture cue. This result is harder to interpret as it is inconsistent with previous studies. In a series of experiments Malcolm and Henderson (2009, 2010) found that the initial saccade latency was significantly affected by stimulus onset asynchrony, but unaffected by cue type. This was interpreted as the initial saccade latency reflecting a general preparatory process, rather than a result of a cue manipulation. In order to determine if this result is a legitimate effect, Experiment 5 will see if Experiment 4's finding replicates or not.

Experiment 5 will also deal with a potential unaccounted for variable in Experiment 4. In Experiment 4 colour information was removed by changing picture cues to grey-scale. This does not actually remove colour information, so much as changes it. As such, Experiment 4's participants may not have ignored colour information of the cue, but instead generated a grey target template and tried to use the grey colour to guide search. In other words, despite being aware of the conditions in the experiment, and knowing that if they saw a grey cue it meant that the target could be in any colour, participants inadvertently assumed the target object was grey. This would mean that as opposed to the visual system guiding search with every feature but colour in the grey-scale condition, participants might have been searching for grey objects to fixate in the periphery, increasing the number of regions visited, and trying to match fixated regions with grey objects, increasing scanning fixation durations and verification time. Wolfe et al. (2004) found that if they gave participants picture cues that indicated the target type, but did not give away any of the specific visual details (that is a picture of another object from the same category; for example, if the target is a dog the cue could be a picture of Boston Terrier but the target dog in the display could be a German Shepherd) their respective search times increased to the same rate as word cues. The authors suggested that this meant that the visual system did not simply prefer picture cues, but the specific low-level visual information about the target. The grey-scaled cues used in Experiment 4 did not differ from the target in visual detail as greatly as the same-category-different-type picture cues used by Wolfe and colleagues; however, the possibility remains that participants in Experiment 4 were inadvertently misled by, rather than ignorant to, the target's colour.

In order to be certain that colour is such a dominant feature and also of the specific role colour plays in search, a second experiment was run. The paradigm remained the same, but instead of seeing which search processes suffered when colour information was the only feature removed, Experiment 5 investigated which search processes were facilitated when colour was the only visual feature available. Therefore, in Experiment 5, the grey-scale picture cue condition was replaced with a word plus colour cue condition. This gave participants a semantic cue plus a representative colour (Figure 9).

Figure 9 Examples of the three cue conditions used in Experiment 5. Note that the colour picture cue would always appear the same size as it would in the scene.

Colour swatches were cut out sections of the target object. Colour swatches were used so as to give participants an estimation of the target's predominant colour. Although this will inevitably carry different variations of a particular colour, it would be impossible to show one colour that was entirely representative of a real-world object. Real-world target objects are rarely one solid colour: the apple in your fruit basket may be considered red, but there are no doubt inconsistencies in that colour. Even objects that are one colour (e.g., a traffic cone that is a consistent orange hue) will not be perceived as a consistent due to inconsistencies of lighting. By giving participants a colour swatch they are provided with the colour gist of the target object.

If colour facilitates real-world target template search as much as Experiment 4 suggests, and the same behaviours as Experiment 4 suggests, then in Experiment 5 the word plus colour cue condition should lead to shorter total trial durations than the word cue condition. More specifically, every eye movement behaviour that was significantly hindered by the removal of colour information in Experiment 4 (e.g., longer fixations, more regions fixated) should now be significantly facilitated when compared with the respective behaviour in the word cue condition in Experiment 5. If colour is a dominant factor in the scanning epoch, particularly in the weighting of the activation map, then Experiment 5 should find that performance in the word plus colour cue condition should be significantly better than the word cue condition, but not different to the colour picture cue condition. That is because if colour information plays such a dominant role in search that its removal hinders search efficiency to the rate of a semantic cue, then colour information is all that should be needed to improve a semantic cue's ability to guide search to near optimal levels (the level of an exact picture cue). Similarly, if colour is as dominant a guidance feature as Experiment 4 suggests, then Experiment 5 should find that the word plus colour cue condition should facilitate the process of matching the fixated region to the target template and making an accept/reject decision.

Experiment 5

Methods

Participants. 18 participants were run. All participants gave informed consent in accordance with the institutional review boards of University of Edinburgh. None of the participants had taken part in Experiment 4.

Stimulus Material. Sixty-three photographs of real-world scenes from a variety of categories (indoor and outdoor, natural and man-made) were used as stimuli. In contrast to Experiment 4, in Experiment 5 a target object was added to every scene (Figure 9). All target objects had one predominant colour as determined by initial pilot testing. Target objects were selected from Hemera Images database (Hemera Technologies Inc., Gatineau, Canada), modified and then placed into scenes with Adobe Photoshop CS (Adobe, San Jose, CA). This was necessary to acquire a colour swatch in the word plus colour cue condition. In the word plus colour cue condition, a 100x100 pixel copy of the target object was taken from an area that contained a predominant colour of the object. Hemera Objects tend to be very large; this allowed a 100x100 pixel region of the target object to be copied and then the object shrunk to a size that suited the image. The 100x100 pixel region, or swatch, was then placed in a box under a word cue. Words were written in 72 point font subtending 2.14 degrees in height, above the swatch. In the word cue condition the cues were exactly the same as the word plus colour cue condition, except that the box beneath the word was empty. Colour picture cues were created simply by copying the target object from the scene and pasting it into a blank background.

Procedure. The procedure was the same as Experiment 4. There were 63 trials, 21 for each condition. Each participant was given 9 practice trials prior to testing, with 3 trials from each condition.

Results

Eighteen participants' data were analysed. The mean accuracy was 93%. Only correct trials were analysed. Criteria for a correct trial were the same as Experiment 4, except now trials over 5500ms were removed from analysis. Similar to Experiment 4, trials where a participant fixated the target, moved off the target for one fixation, and returned the next fixation were scored as correct. This occurred on 2.4% of correct trials.

Analysis of the Overall Trial Duration and Its Search Epochs

The results show a main effect of cue type on total trial duration, $F(2, 34)=51.569$, $MS_E=11467.268$, $p<0.001$ (Table 5). A planned contrast revealed that trials with a word plus colour cue had shorter total trial durations than trials with word cues alone, $F(1, 17)=62.935$, $MS_E= 20027.884$, $p<0.001$. Searches cued by colour picture were shorter than those cued by a word plus colour cue, $F(1,17)=4.504$, $MS_E= 27042.521$, $p=0.049$.

	Word Cue		Word + Colour Swatch Cue		Colour-Picture Cue	
	Mean	**SE**	**Mean**	**SE**	**Mean**	**SE**
Total Trial Duration	1450.16	51.53	1185.54	40.91	1103.28	41.46
Search Initiation Time	236.49	7.11	227.30	6.72	225.97	7.64
Scanning Time	695.59	30.50	487.52	23.62	465.07	26.82
Regions Visited	2.52	0.09	2.11	0.09	2.02	0.08
Fixation Durations	184.06	5.04	175.30	5.85	164.97	5.73
Verification Time	518.29	35.40	470.28	33.32	413.51	34.71

Table 5 Results from Experiment 5. All means are in milliseconds except for Regions Visited which is measured in the number of regions visited on the display screen (maximum of 48).

The three search epochs were then examined. There was no main effect of cue type on initial saccade latency, $F(2, 32)=2.361$, $MS_E=250.078$, $p=0.110$. There was a main effect of cue type on scanning time, $F(2, 34)=33.179$, $MS_E=8764.578$, $p<0.001$. A planned contrast showed that the word plus colour cue condition had shorter scanning epochs than the word cue condition, $F(1, 17)=47.313$, $MS_E=16470.468$, $p<0.001$. However, there was no significant difference in scanning epoch between the word plus colour cue and the colour picture cue conditions: $F<1$. This indicates that participants did not locate the target significantly faster when given an exact visual cue than when given just the target's colour information.

There was a main effect of cue type on verification time, $F(2, 34)=40.730$, $MS_E=1215.712$, $p<0.001$. A planned contrast showed that word plus colour cues had shorter verification epochs than word cues alone, $F(1, 17)=12.872$, $MS_E=3222.750$, $p=0.002$. Colour picture cues also led to shorter verification epochs than words plus colour cues: $F(1, 17)=23.248$, $MS_E=2495.126$, $p<0.001$.

The Underlying Behaviours Affecting Scanning Time

The additional features in a colour picture cue did not improve the scanning process any more than when colour was the lone visual feature available. This supports the suggestion from previous research and Experiment 4 of this study, that a target template's colour information plays a dominant role in guiding search. The two main processes during scanning were then examined: selecting the next saccade destination and processing the fixated object. The results indicated a main effect on the number of regions visited, $F(2, 34)=13.468$, $MS_E=0.096$, $p<0.001$. A planned contrast revealed that participants visited fewer scene regions when cued with a word plus colour than just a word, $F(1, 17)=15.618$, $MS_E=0.201$, $p=0.001$. However, there was no significant difference between the colour picture and word plus colour cue conditions: $F<1$.

There was a main effect of cue type on scanning epoch fixation durations, $F(2, 34)=8.927$, $MS_E=184.243$, $p=0.001$. A planned contrast indicated that word plus colour cues had significantly shorter scanning epoch fixation durations than word cues alone, $F(1, 17)=5.065$, $MS_E=272.970$, $p=0.038$. There was also a trend for colour picture cues to have shorter fixations than word plus colour cues: $F(1, 17)=3.998$, $MS_E=480.593$, $p=0.062$.

Discussion

In Experiment 4 removing colour information from a target template was found to have a significant effect on both the saccadic selection process and the evaluation of fixated regions. However, one potential issue was that though participants were aware that grey-scale cues were not the actual colour of the target object, they still may have inadvertently used the grey features to guide search. The results therefore may have not represented search guided by every potential visual feature except colour, but instead represented search guided by the wrong colour.

Experiment 5 was designed to complement Experiment 4. Experiment 4 investigated what search behaviours were hindered by removing colour information from an otherwise complete target template. Experiment 5 examined what search behaviours were facilitated when colour information was added to an otherwise semantic based target template. Importantly, if colour had as big an effect on both the saccadic selection process and the matching process as Experiment 4 suggests, then the results of Experiment 5 should complement those of Experiment 4.

The results indicated that when participants were only equipped with the target label and a representative swatch of its colour, the scanning epoch was nearly as short as when participants were cued with an exact picture of the target. A closer examination of the scanning epoch indicated that the primary benefit of the colour feature in a target template was in selecting the next saccade destination: participants again visited nearly as few scene regions in the word plus colour cue condition as when given additional visual information in the colour picture cue condition. The word plus colour cue condition was also significantly better than the word cue alone. This supports previous studies suggesting that colour information is dominant in guiding saccades and, as in Experiment 4, demonstrates that this dominance extends to search through real-world scene images. The current study suggests that colour information, when available, is the dominant source of information used by the visual system when guiding eye movements. This is not to say that other features can not be used by the visual system; the saccadic selection process can be guided by size (Hannus et al., 2006; L. G. Williams, 1967), shape (Findlay, 1997; L. G. Williams, 1967), orientation (Hannus et al., 2006; Tavassoli et al., 2009; but see Motter & Belky, 1998) and spatial frequency (Tavassoli et al., 2009). However, it appears that when knowledge of a target's colour is available in a real-world image search task it affects the weighting of the activation map so much that the guidance information provided by all the other features becomes minimal.

When looking at the temporal aspects of search, however, colour does not appear to be as dominant a feature. In Experiment 4, removing colour increased both scanning fixation durations and target gaze durations to the same level as when a word cue was used. However, this could have been due to participants treating grey-scale cues as representative of a grey target rather than a colourless one. In Experiment 5, when colour information was the only type of visual information added to an otherwise semantic cue, both scanning fixations and target verification became significantly shorter than those in the word cue condition. However, the colour picture cue condition produced significantly shorter measures than the word plus colour cue condition. This implies that while the visual system will primarily rely on colour information – at the expense of other available guidance features – to select saccade destinations, the visual system uses more than just colour information to make an accept/reject decision about a fixated region.

There was no effect of cue on search initiation time in Experiment 5. This result supports the hypothesis that initial saccade latency is a general preparatory process not affected by cue type, as suggested in Malcolm and Henderson (2009, 2010).

General Discussion

When looking for a target, observers will generate a template to compare with the display percept and identify regions of high-correlation. Many different features within a template have been shown to facilitate search in some manner. One of the more interesting features that can guide search is colour; previous studies suggest that colour is relied upon more by the visual system than other features when guiding search. Two experiments investigated whether colour was still a dominant feature when searching through real-world scene images and, if so, which specific behaviours were facilitated.

The results clearly indicate that colour information facilitates real-world image search, supporting previous work that had demonstrated a similar facilitation in stimulus arrays. This facilitation can be seen by looking at the total trial duration data and noting that when colour was removed from an otherwise complete picture cue, total trial duration increased; when colour was added to a word cue total trial duration decreased.

These results contrast with those of Schmidt and Zelinsky (2009) who found no significant effect on trial duration when a colour label was added to the cue (and in one case they actually found a trend towards colour information increasing trial durations). However, similar to the study here, they found that when colour information was available search guidance improved.

In the current study, the visual system relied on colour information the most during the scanning epoch when weighting the activation map in order to select the next saccade destination. The addition of colour information to an otherwise semantic cue had such a large facilitation effect on the activation map that performance reached a ceiling level: that is a word plus colour cue provided a level of saccade selection performance that could not be significantly improved by the addition of other features in the colour picture condition. It should be noted that, though not a significant difference, picture cues led to fewer regions of a scene being visited before locating a target than word plus colour cues, suggesting that colour may not play the entire role in target template guidance. However, between the results of Experiment 4 and Experiment 5, there is enough evidence to suggest strongly that target colour is a dominant feature in influencing an activation map. This result adds empirical support to Zelinsky's real-world target template search models (Rao et al., 2002; Zelinsky, 2008) which suggest that the visual system exploits a target template's colour information when locating the position of a target in natural scene images.

The present results not only support these previous studies and models, they also add to the current understanding of how colour information affects the search process. Colour information was found to have a significant effect on the evaluation of fixated regions in which an accept/reject decision was being made. Just as in the saccade destination analysis, the results indicated that removing colour from an otherwise complete picture cue increased the time needed to reject distracters, and adding colour information to a semantic cue shortened the time needed. Similarly, removing colour information significantly increased verification time, while adding colour information to a cue significantly reduced verification time.

However, unlike the saccade selection process, the evaluation process seems to rely comparatively less on colour information. When colour information was the only visual feature available with an otherwise semantic cue, both scanning fixations and verification time became significantly shorter; however, both measures were shorter still when participants were given a colour picture cue. This suggests that colour information plays a more minimal role in the evaluation process than the eye guidance process.

It is somewhat surprising that the visual system would benefit from the availability of extra visual features when making a reject decision during scanning. When rejecting a fixated object as a distracter, only one feature from the fixated object should have to be identified as different from that of the target template. For instance, if the target is a blue mug and the visual system fixates on a black stapler, a visual system with blue as the only available visual feature should be able to reject the fixated black object easily. The current data suggest that if participants had further visual information about the target object available, say shape and size, the stapler would have been rejected more quickly.

One possible explanation for this result relates to the saccade selection process. Since it appears that colour is the feature primarily used by the visual system in selecting regions to fixate, it would seem reasonable to assume that most fixated regions shared a similar colour with that of the target. Since fixated objects were selected for being similar in colour to the target, other features will need to be used to verify whether the object is actually the target or not (e.g., this object has a similar colour to the blue coffee mug, does it have the shape of a mug?). More time will be needed to contrast other features of the fixated object with those of the target template, especially if the target template was generated from a semantic cue which would include no visual details.

Collectively, the results from both experiments suggest that colour is a dominant feature used during the guidance of visual search in real-world scenes. Colour is an important factor in making the comparison between fixated elements and internal representation of the target, but the availability of other visual features significantly shortens this process. Contrast this with the spatial aspect of search where adding colour information to an otherwise semantic cue improves the saccadic decision process so much that the addition of other features does not significantly improve the process.

Refining Our Understanding of Real-World Target Template Guidance

The results supports previous studies indicating that real-world visual search improves the more specific the target template becomes (Malcolm & Henderson, 2009, 2010). However, it seems as if this general statement may need to be refined. It appears that the visual system has a preferred subset of features for guiding search. In particular, when the target template contains colour information the process of selecting the next saccade destination appears to near a ceiling level. After that, increasing the specificity of the target template will provide minimal improvement to this selection process. However, it should not be inferred that available colour information renders all other visual features in a specific template incidental. First, the visual system can use other visual features to guide search in a top-down manner (Findlay, 1997; Hannus et al., 2006; Tavassoli et al., 2009; L. G. Williams, 1967; Wolfe, 1994a); the visual system is therefore obviously flexible in the information it chooses to use in weighting the activation map. Secondly, colour information in a target template seems to reach its peak importance during the saccade selection process. However, in other major search behaviours, such as comparing the fixated region with the target template, colour played a more minimal role meaning that the visual system prefers to include other available features in this process.

When drawn together it can be inferred that the visual system uses different visual features over the course of the task, depending on the current behaviour. This in itself may be why past studies have found that search is facilitated the more specific the target template becomes: the visual system primarily relies on different features over the course of the task. When an exact target template is stored in visual working memory, all the target's features are available for the visual system to utilise when needed during the task. Future studies should look at when other features are utilised during search. For example, target template orientation, both in plane and in depth, has been shown to play an important role in facilitating search (Vickery et al., 2005). However, it is not known when in a search task, and for which behaviours, the visual system relies on orientation information the most. Knowing when the visual system uses orientation, or other features, will help refine our understanding of the target template guided real-world search process.

Chapter 4: Combining Target Template With Scene Context to Guide Real World Scene Search

Real-world search involves guiding eye movements around a scene based on information from the scene percept interacting with task goals. There are three general types of information that can facilitate guidance including image features, scene context and, as Chapters 2 & 3 demonstrated, a target template. The real world, however, is dynamic, and some or all of these information types could be available at any one time. An optimal system would combine simultaneously available information to guide visual search.

Surprisingly there has been a paucity of empirical research into whether the human visual system does, in fact, integrate information types to guide search during a single task[8], though more recent real-world search models have begun to find support for this possibility. In particular, these models demonstrate that their predicted fixation positions bear a closer resemblance to those of humans when they account for more than one type of guidance information (Ehinger et al., 2009; Kanan et al., 2009; Torralba et al., 2006). For example, Torralba et al.'s (2006) *Contextual Guidance Model* combined low-level salience and scene context when guiding search (Figure 10, top). Areas of high salience within a selected global region are given higher weights on an activation map than those that fall outside of the selected global region. These highly activated regions are then selected for fixation. The *Contextual Guidance Model* outperformed a pure salience-driven model in predicting human fixation locations in a search task. In particular, the model predicted the first few human fixation locations with a high degree of accuracy. This increased accuracy in predicting human fixation positions when two information types were combined support the suggestion that humans actively utilise more than one information type to guide search.

[8] Kunar et al.'s (2006) empirical study investigated whether humans actively combine disparate information types to facilitate search. Using object arrays, the authors investigated whether participants integrated both display context and template information. Though they did find some supportive evidence, their results were inconsistent and open to interpretation.

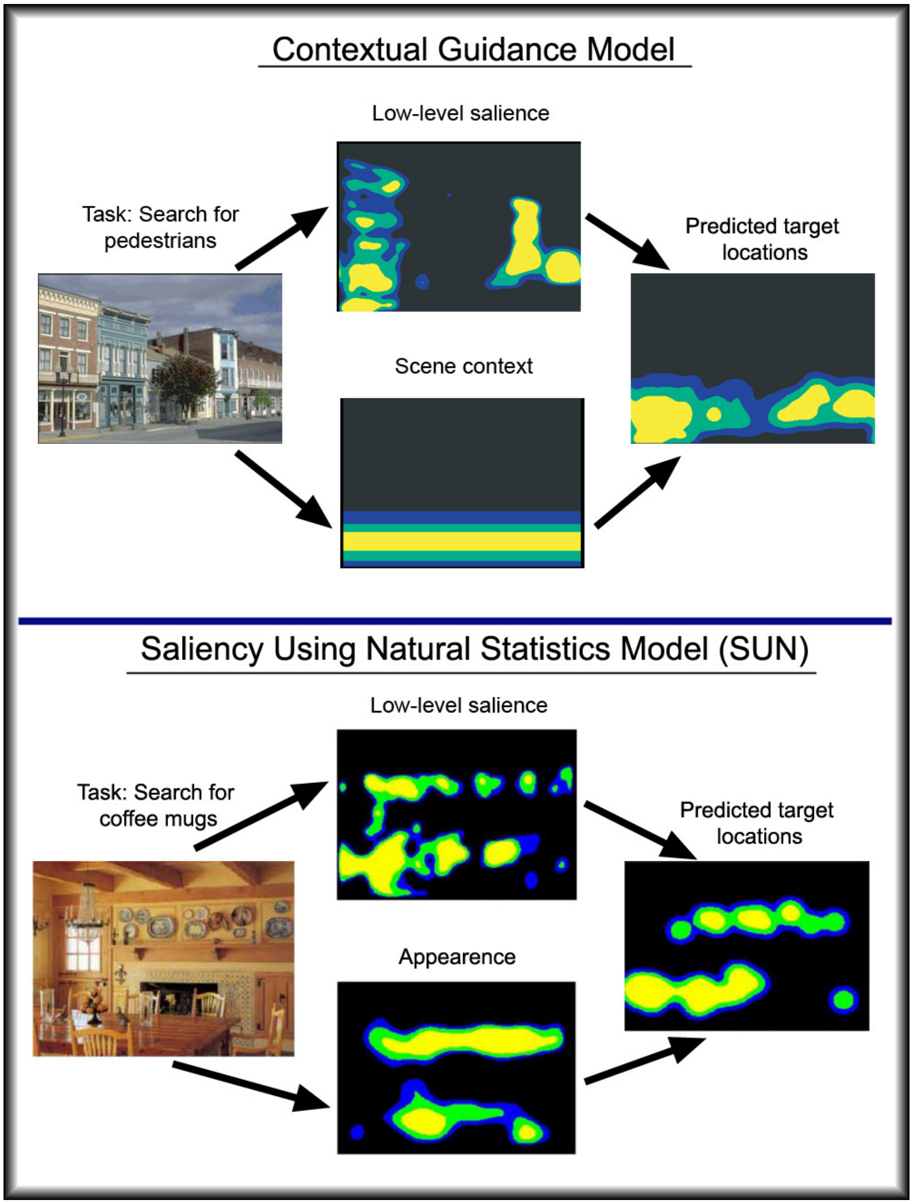

Figure 10 Top: the *Contextual Guidance Model*. The model combines low-level salience with scene context to predict fixation locations. Bottom: the *SUN Model*. The model combines low-level salience with target appearance to predict fixation locations. (Note: parts of these images are taken from Torralba et al. (2006) and Kanan et al. (2009), respectively).

This research has since been updated (Ehinger et al., 2009). Separate models were generated using each of the above three information types, and were again assessed by comparing their predicted fixation locations with those of humans. Ehinger and colleagues found a high degree of agreement in fixation locations among participants, and while models using just one information type out-performed a control, models that combined all three information types accounted for the greatest percentage of human fixation locations.

Similar to the contextual guidance model, Kanan et al.'s (2009) *SUN Model* (Saliency Using Natural statistics) combines top-down and bottom-up information to guide eye movements during real-world image search tasks (Figure 10, bottom). However, unlike the *Contextual Guidance Model*, *SUN* implements target features as the top-down component. *SUN* outperformed a salience-driven model in predicting human fixation positions during real-world image search. The *SUN Model* also slightly outperformed the *Contextual Guidance Model*, though performance was similar overall. The important point is that both models found that combining two sources of guidance significantly improved their abilities to predict human fixation locations, suggesting that humans similarly combine information types to guide search.

Collectively, these results indicate that the visual system does not restrict itself to a single source of information when guiding search. However, there are limitations to what these studies can tell us, particularly with regards to how the visual system combines information. Do these sources of information affect the same search behaviours or different ones? For instance, both scene context (Castelhano & Henderson, 2007; Eckstein et al., 2006; Neider & Zelinsky, 2006) and target template information (Malcolm & Henderson, 2009, see also Chapters 2 & 3) facilitate eye guidance. When both types of information are available in a search task is the benefit on eye guidance additive or superadditive? Or does the system use each information type to facilitate a different search process? For example, if scene context improves guidance to potentially relevant scene regions, might the visual system utilise target template information to facilitate a different search process such as determining if the target is present at each fixation?

In order to answer these questions, Chapter 4's study manipulated the availability of different information types to see how human search behaviour was affected. An eye-tracker was again used to divide search time into three behaviourally defined epochs, allowing fine-grained analysis into which behaviours are affected by the presence of extra guidance information. In particular, recording eye movements revealed if two separate information sources can facilitate the same search behaviours and, if so, in what manner.

Combining Target Template Information With Scene Context

Of the three sets of information available to guide real-world search, Chapter 4 focuses on combining target template and scene context information. This focus represents a departure from previous studies. While previous studies have investigated the relationship between salience and scene context (Torralba et al., 2006), or salience and target template information (Kanan et al., 2009) or all three information types at once (Ehinger et al., 2009), no study has focused specifically on the target template-scene context relationship (Figure 11).

Understanding this particular relationship is important since recent evidence suggests that the visual system relies on target template and scene context information more than low-level image properties when guiding search (Ehinger et al., 2009). This is supported by the growing evidence that the influence of low-level salience information on search guidance is minimal, leaving top-down processing as the dominant processes in real-world search (Ehinger et al., 2009; Einhäuser et al., 2008; Einhäuser et al., 2007; Foulsham & Underwood, 2007; Henderson et al., 2007; Henderson et al., 2009; Tatler et al., 2005; Tatler & Vincent, 2009; Turano et al., 2003; Zelinsky, 2008; Zelinsky et al., 2006). Even the high intersubject agreement of the first few fixations – which was claimed to be a result of the visual system selecting low-level salient areas to fixate (Carmi & Itti, 2006; Parkhurst et al., 2002) – has since been shown to be an artefact, and instead thought to derive from common high-level knowledge strategies (Tatler et al., 2005). Therefore, focusing on how the visual system combines two top-down processes, such as target template and scene context information, will reveal more about how the visual system guides search, than if top-down processing was studied in conjunction with image features.

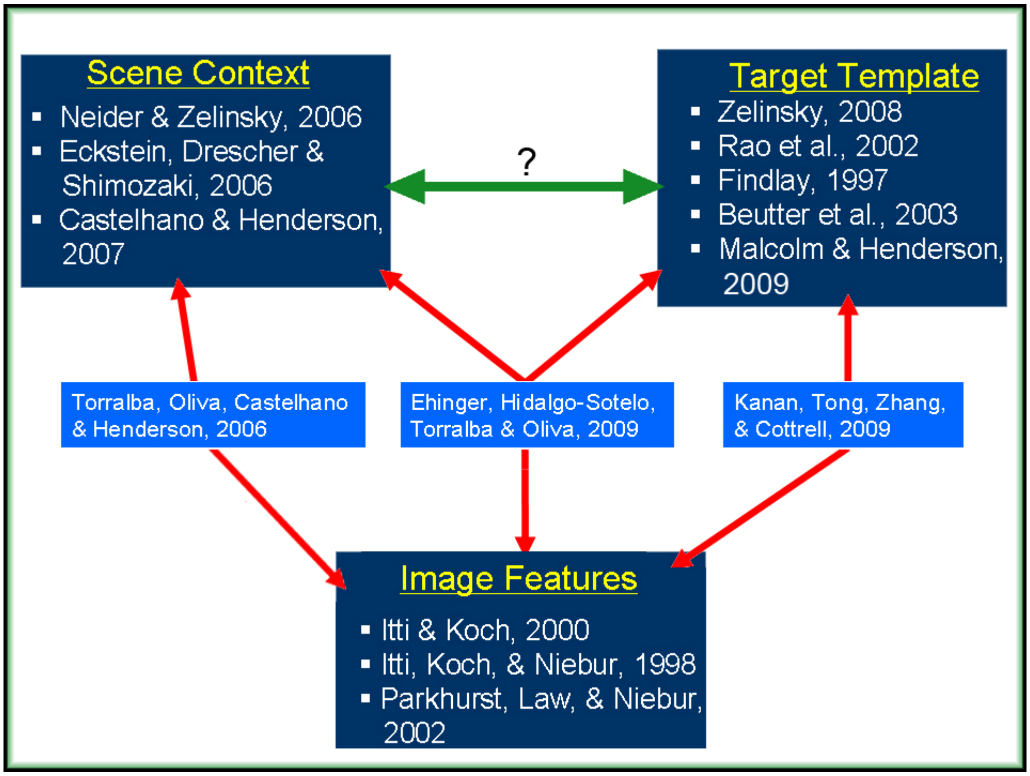

Figure 11 An outline of some of the eye movement guidance research. Individual sources of guidance information have been identified such as image features, scene context and target template information (dark blue boxes). Few studies have examined how these sources of guidance combine to facilitate search (light blue boxes and red arrows). No study has looked specifically at how combining just cognitive knowledge information (i.e., scene context and target template information) facilitates search.

Chapter 4's Study

The present study ran a visual search experiment in which the target could either be cued by an abstract cue (a word) or a specific cue (an exact matching picture of the target), and the target appeared in either a high probability region of a scene (scene regions where the target object was likely to be found: e.g., staplers on desks, ceiling fans on the ceiling, etc.) or a low probability region (scene regions less likely to contain the target object: e.g., staplers on the ceiling, ceiling fans on desks, etc.). As in Chapters 2 & 3, picture cues supported creation of a specific target template that could guide the search process, whereas word cues supported a less specific and more abstract template that would provide less constraint on search guidance. If the target was positioned in a high probability region then scene context would provide information about where the target was located; if the target was positioned in a low probability region, then scene context would not provide information about the target's location.

If integrating target features with scene context provides for a more efficient search strategy than using either source of information alone, then searches with a specific cue and a target located in a high probability region should result in the fastest search times.

Search is also a process that can be broken down into sub-processes that change over the course of time, and analyses that only measure overall search time may miss comparatively more fine-grained behaviours. Therefore, as in Chapters 2 & 3, an eye-tracker was used to divide overall search time into sub-epochs: search initiation time, scanning time and verification time.

Previous results suggest that cue type should not affect search initiation time, but should affect scanning time with more specific cues leading to shorter scanning epochs (Malcolm & Henderson, 2009, see also Chapters 2 & 3). Similarly, previous research has shown that the first saccade in a real-world scene search task tends to land nearer regions that are more likely to contain the target (Castelhano & Henderson, 2007; Eckstein et al., 2006; Neider & Zelinsky, 2006). If scene context benefits search by providing the participant with a probable region within which to search, then there should be shorter scanning times when the target is in a high probability region. In terms of verification time, past research suggests that a more specific target template should facilitate the verification process (Burgess, 1985; Castelhano et al., 2008; Greenhouse & Cohn, 1978; Judy et al., 1995; Malcolm & Henderson, 2009) and that objects located in a high probability region of a scene should take less time to verify (Biederman et al., 1982).

Experiment 6

Method

Participants. Twenty-four participants (sixteen female, ages 19-32, mean age 22.3) gave informed consent in accordance with the institutional review board of the University of Edinburgh. All participants were naïve about the purpose of the study.

Apparatus. Eye movements were recorded using the same eye-tracker as in Chapters 2 & 3.

Stimulus Materials. Fifty-two photographs of real-world scenes from a variety of categories (indoor and outdoor, natural and man-made) were used as stimuli. All images were scaled to 800x600 pixel resolutions and shown in full colour. In each scene, two objects were taken from Hemera Images database (Hemera Technologies Inc., Gatineau, Canada) and then modified and placed into the scene with Adobe Photoshop CS (Adobe, San Jose, CA). Each object appeared in only one scene.

In order to manipulate scene context each scene had two targets assigned to it, only one of which appeared during testing (see Figure 12). Both objects were positioned so as to appear in high probability regions (creating two possible scenes). Targets were then swapped so that they appeared in the reciprocal target's location so that they were now in low probability regions (creating two more scenes).

Eight participants ranked the scenes to determine whether the targets were placed in high and low probability regions as intended. None of these participants took part in the search experiment, and none of them had seen the scene images before. Participants were given a 7-point Likert-like scale and asked to evaluate whether the target was positioned where they would expect to find it in the given scene; 7 for yes, definitely; 1 for no, not at all. Participants were divided into two groups. Each group saw half of all the possible scenes, viewing each scene twice but with a different target each time. A participant never saw the same object at two different locations within the scene. Targets positioned in high probability regions were judged to be in more expected regions of the scene than targets positioned in low probability regions, $t(7)=18.5$, $p<0.001$.

To create the picture cues for the experiment, each target was pasted into a grey background, appearing exactly as it would in the scene. A further 52 corresponding word cues were created that contained only the name of the target object in 72 point font subtending 2.14 degrees in height, centred on a grey background.

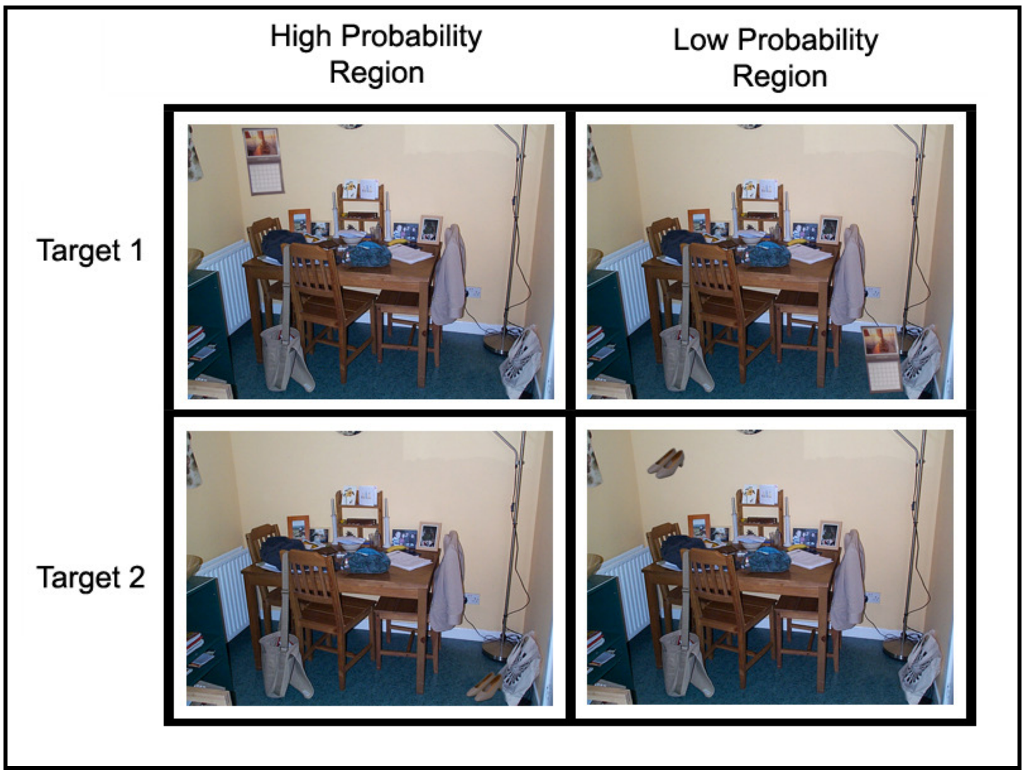

Figure 12 Example of how each scene had two possible targets (here, the calendar and the shoes) in two possible locations (here, on the wall or on the floor) to create four possibilities for each scene. Only one target ever appeared in a scene at a time. During the experiment participants were shown either a word or picture cue of the target, followed by the scene with the target in either a high probability or low probability region. The cued target was always present.

Fifty-two further scenes were added to the experiment as fillers. Filler scenes used an existing object in the scene as the target. Targets in filler scenes were positioned in high-probability locations, meaning that 75% of all the scenes viewed by participants had target objects in high-probability regions. This percentage ensured that participants would recognize scene context as a potential source of guidance throughout the experiment. Half of the filler scenes were cued with words and half with pictures. The eye movements from the filler trials were not analysed.

Procedure. Prior to the experiment, each participant underwent the EyeLink calibration procedure: Eye positions were recorded as participants fixated a series of 9 dots arranged in a square grid extending to 19.25° eccentricity. Calibration was then validated against a second set of 9 dots.

For the experiment, each participant began a trial by fixating a drift correction dot in the middle of the screen (for the purpose of eliminating any tracking error). The experimenter then initiated the trial. A central fixation cross appeared for 400ms followed by a cue identifying the search target for 800ms. The cue was either a word identifying the target or an exactly matching picture of the target. This was followed by a central fixation point lasting another 200ms, making a stimulus onset asynchrony of 1000ms. The search scene then appeared and participants searched for the target object, responding with a button-press as soon as it was located.

There were four possible layouts for each test scene: Target 1 in a high probability region, Target 2 in a high probability region, Target 1 in a low probability region, Target 2 in a low probability region. Participants only saw one of these four types of scene layout for each scene during an experiment, so each of the 52 scenes was seen only once. The four manipulations were rotated through scenes across participants in a Latin Square design. Test scenes and filler scenes were intermixed and occurred in a random order for each participant.

Results

Trials with errors were removed from analysis. Error trials were defined as those in which participants incorrectly identified the target; participants fixated the target, moved away, and returned before correctly identifying it; or the total trial duration exceeded 5500ms. Under these criteria 90.3% of trials were scored as correct. If a participant fixated the target once, moved fixation off the target for one fixation, and then returned to the target on the next fixation, this was accepted as a correct trial. Here, a single fixation deviating away from the target was considered to be the result of a pre-programmed oculomotor command and not due to a decision to attend to a different possible target. This scenario occurred on 4.7% of the correct trials.

A 2x2 repeated measure ANOVA showed a trend for a significant main effect of cue type on accuracy, with word cued trials having marginally higher accuracy scores than picture cued trials (93% and 92% for word cues in high-probability and low-probability regions, respectively; and 89% and 88% for picture cues in high-probability and low-probability regions, respectively), $F(1, 23)=3.702$, $MS_E=1.489$, $p=0.067$. There was no main effect of scene context, nor was there a significant interaction between cue type and scene context, *F*s<1 (Table 6).

	Word Cue, HPR		Word Cue, LPR		Picture Cue, HPR		Picture Cue, LPR	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Total Trial Duration	1375.34	61.08	1765.19	75.41	1092.87	53.92	1351.87	80.12
Search Initiation Time	250.22	10.30	257.86	8.89	241.99	8.94	248.96	8.73
Scanning Time	609.45	41.93	950.35	58.61	445.86	40.36	664.96	56.69
Verification Time	515.53	39.02	556.89	33.30	405.22	26.36	437.95	32.82

Table 6 Results. All means are in millisecond units.

Analysis of Total Trial Duration and the Three Search Epochs

Additional 2x2 repeated-measures ANOVAs, with cue type (word vs. picture) and target location (high probability region vs. low probability region) as factors were conducted on total trial duration, search initiation time, scanning time, and verification time (Table 6).

Both variables produced a main effect on total trial duration, with shorter total trial durations for picture cues and targets located in high probability regions ($F(1, 23)=45.925$, $MS_E=63249.538$, $p<0.001$, and $F(1, 23)=36.587$, $MS_E=69044.168$, $p<0.001$, respectively). There was no significant interaction between cue type and target location, $F(1, 23)=1.413$, $MS_E=72678.914$, $p=0.247$.

There was no main effect of either cue type or target location on search initiation time ($F(1, 23)=1.559$, $MS_E=1130.253$, $p=0.224$, and $F(1, 23)=2.031$, $MS_E=630.416$, $p=0.168$, respectively). There was also no significant interaction between cue type and scene context, $F<1$.

The results of the scanning epoch mirrored those of the total trial duration: both cue type and target location produced a main effect on scanning time with picture cues and targets located in high probability regions generating faster scanning times ($F(1, 23)=24.576$, $MS_E=49213.488$, $p<0.001$, and $F(1, 23)=28.844$, $MS_E=65235.903$, $p<0.001$, respectively). Again, there was no significant interaction between the two factors, $F(1, 23)=1.509$, $MS_E=58968.255$, $p=0.232$.

The verification epoch showed the same pattern of results as the total trial duration and scanning epoch. There were main effects of both cue type and target location on verification time, with picture cues and targets located in high probability regions producing faster verification times ($F(1, 23)=52.641$, $MS_E=5990.367$, $p<0.001$, and $F(1, 23)=8.212$, $MS_E=4011.385$, $p=0.009$, respectively). Again, there was no significant interaction between cue type and scene context, $F<1$.

As stated above, only correct trials were analysed; that is, trials in which the participant fixated the target in the allotted time and pressed the response button. Cases in which participants fixated the target and then saccaded away from it were considered errors and the trial was not analysed, except when the participant saccaded away for exactly one fixation and then immediately returned with the next fixation. In this case the saccade away from the target was treated as a result of a pre-programmed oculomotor command and not due to a decision to reject the target and attend elsewhere. If this scenario occurred during one condition more than another it may have biased the verification time analysis. However, when the data were re-analysed it was found that this was not the case.

The results indicate that both information types facilitate locating and verifying the target object but not initiating the search process. When scene context and a specific target template are both available, they facilitate search in an additive manner.

The Underlying Behaviours Affecting Scanning Time

The results suggest that both cue type and target location have similar effects on the scanning process during search. However, within the scanning epoch there are further sub-processes that the two variables may affect differently. Specifically, there are two decisions occurring during each fixation in the scanning epoch: the visual system must process the fixated object (making a decision about whether it is the target), and when the object is rejected, must decide where to fixate next (Fischer, 1999; Henderson & Ferreira, 1990; van Diepen et al., 1998; Zelinsky, 2008). In order to gauge whether target template and scene context affect these scanning sub-processes in similar or different ways, fixation durations and the number of scene regions fixated were compared as a function of the two variables.

Fixation durations reflect processing time in scene viewing and reading tasks (Henderson & Ferreira, 1990; Henderson & Pierce, 2008; Henderson & Smith, 2009; Rayner, 1998). In previous real-world scene search tasks a more specific template lead to shorter scanning fixations, meaning that fixated distracters were compared with the target template and rejected faster (Malcolm & Henderson, 2009; see also Chapters 2 & 3). This was most likely due to a specific target template having more features to compare against the fixated region: if any one of those features mismatched the fixated region a reject decision could be made quickly. In the current study, if having a more specific template allowed the visual system to process and make a quicker rejection decision at each fixation, then picture cue trials should produce shorter scanning fixations.

The number of scene regions fixated during scanning reflects how spatially selective or distributed search was across conditions. If the visual system capitalizes on scene context about likely target locations, scanning eye movements should be more selective when the target is in an expected location, and fewer regions should be fixated. If the visual system uses a target template to select where to fixate in real-world images, then fewer regions should be visited in the picture cue condition.

It is also worth noting that scanning time showed an additive effect when both information types were available to the participant. There are three possible ways that both information types could affect the scanning sub-processes to produce this effect. First, target template and scene context information could affect both scanning sub-processes; if so, then the availability of both information types should produce an additive effect on both the regions visited and the fixation duration measures. Second, target template and scene context information could affect only one scanning sub-process; if so, then the availability of both information types should produce an additive effect on either the region visited or fixation duration measures. Third, the visual system may utilise one type of information to facilitate one scanning behaviour, and the other type to affect the other scanning behaviour; if so, then there should be a main effect of one information type on one process, and the other information type on the other process (e.g., scene context affects the number of regions visited, but has no affect on the fixation durations, whereas target template information has the opposite effect).

We found that picture cues produced shorter fixation durations than word cues during scanning, $F(1, 23)=46.106$, $MS_E=262.793$, $p<0.001$, replicating Malcolm and Henderson (2009). Targets located in a high probability region also produced shorter fixation durations during scanning, $F(1, 23)=22.785$, $MS_E=335.030$, $p<0.001$. Finally, cue type and target location did not interact, $F<1$ (Table 7). The visual system can process fixated regions faster when both information types are available, with both processes facilitating search in an additive manner.

	Word Cue, HPR		Word Cue, LPR		Picture Cue, HPR		Picture Cue, LPR	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Scanning Time Fixation Duration	188.71	6.15	208.28	6.97	167.98	6.65	184.07	6.11
Scanning Time Regions Visited	2.35	0.14	3.03	0.16	1.94	0.13	2.83	0.17
Verification Time Fixation Duration	247.47	9.58	261.05	11.23	224.12	11.62	241.19	13.53
Verification Time Fixation Count	2.06	0.11	1.96	0.07	1.89	0.11	1.92	0.14

Table 7 All means are in millisecond units except for Regions Visited, which is measured in the number of regions visited on the display screen (maximum 48), and Fixation Count.

Picture cues led to a more selective distribution of attention over the scene during scanning, with fewer scene regions fixated, $F(1, 23)=8.015$, $MS_E=0.280$, $p=0.009$, replicating previous results (Chapters 2 & 3; Malcolm & Henderson, 2009). Similarly, attention was more narrowly distributed when the target appeared in the high probability compared to the low probability region, $F(1, 23)=27.616$, $MS_E=0.533$, $p<0.001$. Again, there was no interaction between cue type and location context, $F<1$. The visual system is more selective in deciding where to fixate when both information types are available, with both processes facilitating this decision in an additive manner.

The results indicate that the additive effect found in scanning time when both information types were available was a result of each scanning sub-process being affected in an additive manner.

The main effect of target location on fixation duration is surprising as it implies the speed with which distracters are rejected is dependent on the target's location, even before the target is located. A potential explanation might come from participants finding the target faster when it is in a high-probability region. Trials in this condition have fewer scanning fixations and average fixation durations are known to increase during a trial (Antes, 1974; Buswell, 1935; Castelhano et al., 2009; Tatler & Vincent, 2008; Unema et al., 2005). When the target is located in a high-probability region there will be fewer of these later, longer fixations (since the target is found faster). This could explain why the mean fixation duration was shorter.

If the effect of target location on fixation duration is solely a result of these later fixations, then there should be no difference between conditions if only the early fixations are compared. The mean duration of these early fixations would not be influenced by the increased duration of later fixations. However, if another factor is causing the effect, then there should be a main effect of target location on scanning fixation durations even during these early fixations. In order to investigate this possibility the analysis focused on the first two scanning fixations. This corresponded to the mean fixation count in the shortest condition (trials with a picture cue and the target in high probability location had a mean of 2.2 scanning fixations). There was a main effect of cue with picture cues producing shorter fixations, $F(1, 23)=38.291$, $MS_E=374.098$, $p<0.001$, and a main effect of target location with targets in high probability regions producing shorter fixations, $F(1, 23)=17.774$, $MS_E=203.572$, $p<0.001$. There was no interaction between the variables, $F(1, 23)=1.001$, $MS_E=248.644$, $p=0.327$. The results do not rule out the possibility that the number of fixations affects the mean fixation duration, but they do rule out the possibility that the number of fixations was the only cause of the main effect of target location. This suggests that there is another factor stemming from the target's location that influences fixation durations.

Another possible explanation for the influence of target location on scanning fixation duration relates to the processes occurring during scanning fixations: The fixated region is evaluated and the subsequent saccade is planned (van Diepen et al., 1998). Given that the mean scanning fixation count was 2.2 in the picture cue/high probability region condition, it is possible that in some of these trials the target was identified in the periphery. This would happen more often in conditions where the target was positioned in highly probable region since in these cases initial saccades would probably have been directed toward the appropriate location. In these trials the verification of the target object may have actually begun prior to the target being fixated. Thus, the final scanning fixation in this particular condition may not reflect typical scanning processes. In order to ensure that the analysis of the mean scanning fixation duration did not include any of the verification process, the mean durations across trial types were re-analysed, but now without including the final scanning fixation.

When the mean scanning fixation durations were analysed, excluding the final scanning fixation, a main effect of cue was found with picture cues producing shorter fixations, $F(1, 23)=11.702$, $MS_E=521.677$, $p=0.002$, and a main effect of target location was found with targets in high probability regions producing shorter fixations, $F(1, 23)=5.352$, $MS_E=815.871$, $p=0.030$. There was no interaction between the variables, $F<1$. These results indicate that the shorter scanning fixations found in trials with the target positioned in a high-probability region were not a result of participants identifying the target in the visual periphery just prior to landing on it. Of course, it could be that in some cases the target was identified two fixations or more prior to its direct fixation, but given how quickly the targets were often found in the present study, it is not possible to test this hypothesis with the current data.

The Underlying Behaviours Affecting Verification Time

Similar to the scanning epoch, the verification epoch was shorter following more specific cues and with targets located in high probability regions. As with the scanning epoch, the verification epoch can be analysed with more detailed measures, including fixation durations and fixation count. Both fixation duration and fixation count reflect how easily the participant can process the target object once fixated. Both verification fixation durations and verification fixation count were therefore examined. For both analyses, the data included only those fixations that were on the target. Though trials in which the participant fixated the target, saccaded away for one fixation and then immediately returned were accepted as correct, the fixations occurring outside the target were not included in these analyses.

Verification fixation durations were shorter following picture cues, $F(1, 23)=12.014$, $MS_E=932.595$, $p=0.002$, and with targets located in high probability regions, $F(1, 23)=8.103$, $MS_E=695.498$, $p=0.009$. There was no interaction between cue type and location context, $F<1$ (Table 7).

Fixation count was not significantly affected by cue type, $F(1, 23)=2.406$, $MS_E=0.113$, $p=0.135$, or target location, $F<1$. There was no significant interaction between cue type and scene context, $F<1$.

Discussion

The present study asked whether the visual system can combine target template with scene context information to facilitate a real-world search task and, if so, how these information sources are combined and which underlying processes are affected.

The results indicated that either information type, by itself, reduces total trial duration; when both forms of top-down information are available, total trial duration is reduced additively. The results thus indicate that the visual system will actively use multiple sources of top-down information to facilitate a search process. The results also indicate that the visual system treats scene context and target template information independently.

The facilitation of scene search by contextual constraint and template specificity was the result of facilitation during two specific search epochs: scanning and verification. Search initiation time, unlike the other two epochs, was unaffected by either variable. When the scanning epoch was further analysed by examining the durations and spatial distributions of fixations during scanning, it was found that target template and scene context produced main effects on each measure. When both information types were available, they influenced both scanning measures additively. This result suggests that scene context and target template information facilitate similar processes during the scanning epoch. Finally, scene context and target template information also affect the same verification behaviours. Each information type affected verification fixation durations additively when both were available. Neither affected verification fixation count.

The similarity between the search behaviours affected by the two information types is surprising in one instance. Scanning fixation durations reflect the processing time needed to compare a fixated region to a target template and make a reject decision. This process is facilitated when a specific target template is available, presumably because the specific target template has visual features which can be quickly compared to the fixated region (Malcolm & Henderson, 2009). However, this process was also found to be facilitated by scene context. This is surprising as it implies the speed with which distracters are rejected is dependent on the target's location, *prior* to the target being fixated. This result seems at odds with the type of information scene context provides: scene context information only indicates which region of a scene a target should be located in, but not the properties of the target. Since scene context does not generate any internal visual properties of the target, what then causes fixated distracters to be rejected faster when the target is located in a high probability region? One possibility is that this effect is a consequence of participants finding targets faster when they are located in high-probability regions. Since early fixations are short and later fixations are longer (Antes, 1974; Buswell, 1935; Castelhano et al., 2009; Tatler & Vincent, 2008; Unema et al., 2005), and since there would be fewer of these later, longer fixations in the high-probability condition (since the target is found quicker), the average fixation duration would be shorter. However, when the analysis focused on just the mean duration of the first two scanning fixations – which were not influenced by the increased duration of later fixations – there was still an effect of target location. A second possibility is that participants sometimes identified the target prior to fixating it. This is particularly likely when the target was in a high-probability region since the initial saccade would be directed towards the appropriate region. In this condition the last scanning fixation may not actually reflect the typical scanning processes. However, when the mean scanning fixation durations were analysed again, but this time without including the final scanning fixation (which may have included some of the verification process), there was still an effect of target location. These two analyses do not entirely rule out the possibility that the effect of target location was due to later fixations or initiation of the verification process, but they do reduce their plausibility.

Combining Information Types in Real-World Visual Search Models and Future Questions

The results suggest that any real-world search model will not only have to identify potential sources of top-down information available in a scene, but also how the visual system combines these sources of information to guide fixation placement and duration.

Focusing on the fixation placement analysis, it is clear that both information types are utilised during the scanning epoch. Scene context provides global information, directing the eyes toward high probability regions of the scene, while the target template distinguishes probable targets from distracters. A future question when examining fixation placement is whether multiple sources of information are used simultaneously during scanning (the visual system selects local features that match the target within a probable global region) or whether the visual system alternates between the two (the visual system uses scene context to select a probable region and then, once there, target template information to select a target-similar object). Scene context and target template information provide different sets of coordinates in suggesting where the target is likely to occur, implying functionally disparate goals in saccadic targeting. There is also some evidence to suggest that scene context may direct search initially without the use of a template (Castelhano & Henderson, 2007; Eckstein et al., 2006; Ehinger et al., 2009; but see Kanan et al., 2009). It could be the case that instead of the visual system using both information types concurrently to locate a target, the visual system utilises scene context and target template information for different purposes. If the visual system does not rely on both information types equally throughout a search task, then identifying when and why the visual system relies on each information type during a task will help reveal the nature of the visual system's search strategy.

Previous analysis of saccade amplitudes, both in image viewing (Tatler & Vincent, 2008; Velichkovsky, Joos, Helmert, & Pannasch, 2005) and real-world viewing (Hayhoe, 2000; Hayhoe & Ballard, 2005; Land & Hayhoe, 2001; Land et al., 1999), suggest that saccades can be divided into those directed toward near-by and far-away targets. Considering that scene context directs the eyes toward selected global regions, it could be that context is used intermittently by the visual system for the purpose of selecting new scene regions to search. Between these saccades, the visual system would search for a target within a region: here scene context would be of minimal use while a target template would be very useful. In this way the visual system can guide search optimally while alternating between the two information types. Whether the visual system utilises scene context and target template information simultaneously or sequentially to guide search is, unfortunately, not evident in the current data since complex scenes were used with high probability and low probability regions that were not easily divisible by an obvious boundary. A saccade could move in the direction of a high probability region, but land on a low probability region, masking the goal of the saccade (e.g., if the participant were searching for a plate in a restaurant, they might saccade downward toward a table in the foreground, but land on the floor in a gap between two tables. Does that mean that the participant was directing their eyes toward the table and came up short, or that the participant was looking for the plate on the floor?). Issues like this make it difficult to determine which particular region a participant was intending to fixate in the current study. Regions must be easily defined if researchers are to separate fixations by region and begin gauging the visual system's underlying goals.

A second issue to be resolved in a future real-world search model is fixation durations. Most current attention models for real-world scenes concentrate solely on where the participant will fixate next. However, as demonstrated previously and supported here, fixation durations in scene perception tasks are not constant: weighing fixated regions by their relative durations changes the attention map of a participant dramatically (Henderson, 2007). The current data not only confirm that fixations differ over the course of a scene search task, but that they differ depending on what top-down information is available, with more information leading to shorter fixations. Future models will have to ascertain how these sources are added together to facilitate quick accept/reject decisions. For the reasons listed above it is difficult in the current study to determine whether multiple sources of top-down information are combined at each fixation in the scanning epoch to reduce processing time, or if the visual system relies on only one source at a time to process a region, alternating which source to use throughout the epoch. In the verification epoch, however, evidence suggests that multiple information sources can be used simultaneously to reduce processing time. There were very few verification fixations (ranging from means of 1.9-2.1 across conditions) and target template and scene context had an additive effect on fixation durations. This suggests that the two sources simultaneously facilitate the process of evaluating the target, reducing verification time accordingly.

Conclusion

When available, the visual system actively uses multiple types of top-down information to facilitate search. This facilitation occurs during the scanning epoch, particularly in the processing of the fixated region and selecting the next fixation position. Both information sources similarly reduce verification time by reducing fixation durations. Each behaviour was facilitated additively when both information types were available, suggesting that the two information types were treated independently by the visual system. Future studies will be needed to probe whether the visual system uses scene context and target template information simultaneously to direct saccades during the scanning epoch, or alternates between the two.

Chapter 5: Discussion

The visual system can guide search through real-world scenes by using different types of guidance information such as low-level image features, scene context and memory. However, guidance by target knowledge remains an under-researched aspect of real-world search. In order to fully understand how the visual system guides real-world search, target template guidance has to be demonstrated in scenes and its contributions to real-world search recognised. Three empirical studies in this thesis have outlined the role target knowledge plays in guiding search.

Chapter 2 investigated if a target template could guide search through real-world scenes and, if so, which exact search behaviours were facilitated by target knowledge. Participants were given real-world search tasks while both cue and SOA were manipulated. Analysis of the resulting trials was segmented into three independent, behaviourally-defined epochs: search initiation time, scanning time and verification time. Three experiments confirmed that picture cues reduce total trial duration, indicating that target knowledge is one method by which the visual system can facilitate real-world search. This effect replicates with real-world scenes previous object array studies that found a facilitation effect when a target template was available (e.g., Wolfe, 1994a; Wolfe et al., 1989). When the three search epochs were analysed it was found that shorter total trial duration in the picture cue condition was due to facilitation of both the scanning and verification epochs. That is, the more specific a target template became the faster a participant located and verified the target. Within the scanning epoch, a specific target template was consistently found to reduce the number of scene regions visited and the mean fixation duration. This was taken as evidence that the visual features of a specific template affects the weighting of the activation map by increasing activation for regions with target-similar features, and facilitates the evaluation process made at each fixation when making an accept/reject decision.

None of the studies found evidence to suggest that target knowledge affected the time to begin the search process. This contrasts with more recent target template studies that found increasing SOA led to increasing search times if the participant had a specific target template, and decreasing search times if the participant had an abstract cue (Vickery et al., 2005; Wolfe et al., 2004). Authors from both studies interpreted these results as evidence that the visual system sets up a specific target template faster than an abstract one before initiating search, and that these effects on the set up time affect the overall search time. However, not only did Chapter 2's experiments fail to replicate this interaction between cue type and SOA on total trial duration, no interaction was found when analysis was focused on search initiation time, a more sensitive target template set up time measure.

This is not to say that the current results rule out the possibility that template set up time is variable. Vickery et al. (2005) and Wolfe et al. (2004) still may be correct in their suggestion. An obvious difference between those studies and the current one is that both previous studies used objects arrays whereas the emphasis in this thesis was real-world search. Object arrays contain stimuli with no established spatial relationships between each other, and the only top-down information available to guide search through novel arrays is target knowledge. Thus variances in the target template set up time could easily affect overall search time. By comparison, a scene is a semantically and spatially coherent view of the real world that contains other types of top-down information able to guide search. A potential reason for not finding the cue type/SOA interaction here is that when a participant knows that other information will be available in the scene to guide search, they may devote fewer cognitive resources to creating a detailed target template leaving the template less susceptible to changes in SOA. The important point is that Chapter 2's study did not rule out the possibility that target template set up time varies according to specificity. What the current results suggest is that when there are other types of guidance information available in a display percept, such as with a scene, search will initiate regardless of the template's state. Real-world search models support this suggestion, finding evidence that the visual system relies on scene context information more than target knowledge when guiding search, though target knowledge still plays a significant role (Ehinger et al., 2009).

Chapter 3 then took the basic findings of Chapter 2 – that target knowledge facilitates scanning and verification processes during scene search – and investigated if colour is relied upon more by the visual system than other features when guiding search. Previous studies into object array search have demonstrated that knowledge of a target's colour affects gaze control more than size, shape and orientation (Hannus et al., 2006; Motter & Belky, 1998; D. E. Williams & Reingold, 2001; L. G. Williams, 1967; but see Findlay, 1997). If the same is true in real-world search, then the effect of template specificity found in Chapter 2 may be refined simply to knowledge of the target's colour. By using an eye-tracker, the exact search behaviours affected by colour information can be identified.

Using the same paradigm as in Chapter 2, two experiments manipulated the availability of colour information and compared the resulting total trial duration and individual epochs with those from word and picture cues trials, allowing analysis into which specific behaviours were facilitated. The results clearly indicated that colour information facilitates real-world image search. When colour information was added to a word cue total trial duration decreased; when colour information was removed from an otherwise complete picture cue total trial duration increased. Colour information had its greatest effect on scanning time, especially the region visited analysis. When a colour swatch was added to a word cue, the region visited analysis improved so much that it was no longer significantly worse than the picture cue condition. This suggests that the visual system relies heavily on a target's colour when selecting target-similar regions in the scene to fixate. It should be noted though that, even though not significant, picture cues required even fewer regions of a scene to be visited before a target was landed upon than word plus colour cues; this suggests that the visual system uses more than just a picture's colour information to guide search. However, there is enough evidence between the two experiments to suggest that colour is a dominant feature in influencing the activation map. These results suggest therefore that target template facilitation of eye movement guidance in Chapter 2 were due largely, though not completely, to the picture cue's colour.

Although colour also clearly facilitated the evaluation process at fixation – adding colour information shortened scanning fixations and the verification epoch while removing colour had the opposite effect – this particular behaviour did not seem to be as influenced by colour manipulations. When colour information was added to a semantic cue the resulting evaluation times became significantly faster, but were still significantly slower than when a picture cue was used. It is therefore evident that the use of other target features improves the evaluation process. This reliance on further template features during the evaluation process could be a direct result of the dominance colour information appears to have in weighting the activation map. If colour plays a large role in selecting fixation sites, it would suggest that most fixations are on regions with similar colours to that of the target. If the site of fixation has a target-similar colour, other features (shape, size, etc.) will be needed to verify whether the object is the target. Support for this hypothesis, however, would require further testing.

Collectively these results suggest that the visual system prefers using different features for different search behaviours. The spatial selection process in particular seems to rely on a particular subset of features in which colour appears to be the most dominant.

With target template guidance now established and the exact methods of facilitation to search understood, Chapter 4 began to examine how target template guidance fit into the entire real-world search process. The study specifically looked at how both target template and scene context information, two cognitive knowledge factors, combined to guide search and which underlying processes were affected. It was found that when either information type was available, search was facilitated. When both types of information were available, search was facilitated additively. The results thus suggest that the visual system actively combines scene context and target knowledge, but that these information types are processed independently by the visual system. When specific search behaviours were examined both information types were found to facilitate the scanning and verification epochs, though not search initiation. Within the scanning epoch, both cognitive knowledge factors affected both the saccade selection and fixation evaluation processes. Collectively, these results suggest that scene context and target template information facilitate similar processes during the scanning epoch. In every case, the availability of both factors was found to facilitate the specific behaviour in an additive manner, again suggesting that each information type was processed independently before facilitating the current search behaviour.

A future study should address whether these types of information are utilised simultaneously or sequentially by the visual system when guiding search (see Future Directions, below, for an in-depth discussion). Analysis of the verification epoch in Chapter 4's study suggests that both cognitive knowledge factors were used simultaneously by the visual system. This epoch, however, occurred only after extended viewing, and only applies to the evaluation process. The visual system's timeframe in utilising two separate information sources to locate a target, especially during the initial stages of search, remains a topic open for investigation.

Together, the three empirical studies of this thesis indicate that target knowledge is a form of real-world guidance that the visual system can rely upon to help locate a target object. The thesis extended results from previous object array search studies that measured both RT (Egeth et al., 1984; Kunar et al., 2006; Vickery et al., 2005; Wolfe, 1994a, 1998; Wolfe & Horowitz, 2004; Wolfe et al., 2004) and eye movements (Castelhano et al., 2008; Chen & Zelinsky, 2006; Findlay, 1997; Luria & Strauss, 1975; Najemnik & Geisler, 2005, 2008; Rajashekar et al., 2006; Tavassoli et al., 2009; L. G. Williams, 1967) into ecologically valid displays, and discovered which exact search behaviours were facilitated and which were independent of target knowledge. By highlighting the dominant effect colour information has during the search guidance process, the study both indicates that the visual system prefers certain features and that the identity of these preferred features vary depending on the current search behaviour. Finally, the thesis offers some of the first evidence as to how target knowledge combines with other types of real-world guidance to facilitate search. When both target and contextual knowledge sources were available to the visual system simultaneously each search behaviour was facilitated additively, suggesting that the visual system processes target template and scene context information independently.

By revealing the nature of target template guidance the present thesis not only contributes to our understanding of the real-world search process, but also reveals the nature of some of the visual system's processing mechanisms.

The Role of Target Template Guidance in Real-World Search

The results from the three studies of this thesis clearly indicate that target knowledge facilitates real-world search, providing empirical support to previous template guided real-world search models (Ehinger et al., 2009; Kanan et al., 2009; Rao et al., 2002; Zelinsky, 2008). Our current understanding of real-world search must therefore be updated to accommodate this guidance factor. The question now becomes what function does target template guidance provide the visual system? The answer appears to be one of flexibility.

As discussed earlier, search appears to be dominated by top-down processes (Ehinger et al., 2009; Einhäuser et al., 2008; Foulsham & Underwood, 2007; Henderson et al., 2009; Turano et al., 2003; Zelinsky, 2008; Zelinsky et al., 2006). In novel scenes, scene context is a form of top-down information that has been empirically shown to guide search (Eckstein et al., 2007; Neider & Zelinsky, 2006). In some circumstances scene context can provide most of the guidance needed to locate a target object in a scene. For example, in Figure 13 the target is the banana. In the top photo contextual constraints will suggest that the target should be on the table, thereby dramatically reducing the area that needs to be searched from the entire environment to a comparatively small region. Although helpful to the process, a target template would seem to be of secondary importance for search guidance.

Figure 13 An example of how the spatial aspects of a scene can affect the method of guidance used by the visual system. In the top picture when looking for the banana, scene context can be used to guide attention to the table, reducing the search space from the entire scene to a relatively small area. In the bottom picture the same table is now seen up close making it part of the back ground. Scene context is now of minimal use and target knowledge is the more effective means of guidance. This demonstrates that in being able to use target knowledge to guide search, the visual system has made itself much more flexible, and thereby robust, in guiding search through dynamically changing environments.

However, the mobility of humans necessarily makes scenes variable in their spatial scale. A scene can be viewed from far away or close up. In Figure 13, despite the two scenes having largely the same content, the perceived space greatly changes the relative worth of different types of guidance information. In the bottom photo the same table now occupies much of the visual field meaning that contextual information provides minimal guidance to the visual system. Target knowledge now becomes the more efficient method of search, distinguishing distracters from possible target objects.

By being able to guide search with either context or target knowledge the visual system has two top-down forms of guidance at its disposal, each specialising in guidance at a different spatial scale. The visual system can thus guide search effectively through a wide range of environments that humans are likely to encounter.

Updating Existing Models of Target Template Guided Real-World Search

There are a handful of real-world search models available right now that incorporate target knowledge, all of which predict human fixation positions with relatively high accuracy (Ehinger et al., 2009; Kanan et al., 2009; Rao et al., 2002; Zelinsky, 2008). These models share in common the premise that the visual system compares a stored set of features with those of the scene percept, selecting regions of high-correlation for fixation. The current results provide empirical support by demonstrating that search is guided by the visual features of the template and that when target features become uncertain the eye guidance process breaks down.

The results of the current thesis can also help improve the accuracy of future models. First, results from Chapter 3 suggest that not all template features are weighted equally by the visual system. Colour, in particular, seems to be a feature preferred by the visual system when comparing a scene to a target template. Future models could test this by increasing the weight provided by colour information in a target template relative to other features and analyse whether the ensuing search pattern becomes more similar to those of humans.

Secondly, these models only account for target template facilitation in terms of fixation placement, but make no predictions regarding changes to the evaluation process of fixated objects (but see Zelinsky, 2008). The current results support the inclusion of a temporal aspect to future search models and indicate that the processing of fixated objects, whether they are the target or a distracter, is faster when a specific template is available for comparison.

The current results also provide extensions to those models which focus on combining guidance information sources (Ehinger et al., 2009; Kanan et al., 2009; Torralba et al., 2006). These models all find evidence to suggest that humans actively combine information to guide search. Both Ehinger et al. (2009) and Torralba et al. (2006) found that by combining the global restraints of scene context with information from a more local spatial scale, either from target template or image features, the resulting search pattern became more correlated with those of humans. Similarly, Kanan et al. (2009) found that when combining target knowledge and saliency the model often made similar false-positive fixations as humans. However, these models do not suggest how the information is combined to facilitate search. Future models will also have to address how these information types are combined, both spatially in terms of eye guidance and temporally in terms of object evaluation. The results from Chapter 4 suggest that in both instances, the presence of contextual constraints and target knowledge have an additive effect.

Future Directions

The collective results not only increase our understanding of the real-world search process and the visual system in general, they also suggest future directions to take target template guidance research. There are two directions, in particular, which could dramatically increase our understanding of the visual system.

The first is to investigate how the visual system utilises multiple types of information to guide search. Previous models (Ehinger et al., 2009; Kanan et al., 2009; Torralba et al., 2006) and results from object array search studies (Kunar et al., 2006; Wolfe, Reijnen, Van Wert, & Kuzmova, 2009) suggest that participants combine available information types to guide search through scenes, a suggestion empirically supported by the results of Chapter 4. The next question is to understand when each information source is used by the visual system to guide eye movements optimally over the course of a search task. This means understanding whether different sources of information are utilised simultaneously or sequentially by the visual system and, if sequentially, in what order different cognitive guidance factors are used in.

Current integrative search models have not resolved this issue, with one model suggesting that scene context is used initially to guide search followed by target knowledge thereafter (Ehinger et al., 2009), while another suggests that target knowledge is used immediately (Kanan et al., 2009). Complicating the issue is that the models' results are, in fact, fixation placement data, not fixation order. The search models do not predict the temporal order of fixations, only fixation positions. If researchers are to understand the visual system's strategy in using disparate sources of information to guide search, empirical evidence is needed documenting when each information source is used over the course of a task.

Figure 14 An example for an experiment investigating when context and template information are used during search. Each scene would have two easily identifiable regions and two semantically contextually appropriate objects. Here the search target is the boat. By manipulating cue specificity and target location, the resulting saccade direction will reveal how the visual system begins the search process.

In order to begin answering this question, an empirical search study could be run in which every scene had two easily divisible global regions and a single semantically appropriate object inserted into each one (Figure 14). In order to manipulate target knowledge the target object would be cued by either a word or picture. In order to manipulate scene context the target will either be positioned in an appropriate or inappropriate location.

With the global regions now easily identified and cleanly bounded, the time frame in which the visual system utilises scene context and target template information can be determined. If the visual system only uses scene context to direct the first saccade, then the high-probability scene region should be saccaded to independently of the presence of a target or distracter, or even the absence of an object in the high-probability area. As a target template would not be influencing the initial saccade direction, the data pattern should also be the same across the picture and word cue conditions. If the visual system directs the first saccade using scene context and the presence of an object in the high-probability region, but does not utilise a target template, then initial saccades should be directed toward the high-probability region when either the target or distracter object is present there, but not when the high-probable region is empty. Again, as a target template is not being used to guide this initial saccade, the results should be independent of cue type.

However, if the visual system uses both scene context and target template information to direct the first saccade, then in the picture cue condition participants should only direct the initial saccade to the high-probability region when the target object is present. However, in the word cue condition the initial saccade should be directed to the high-probability region when either the target or the distracter object is there.

Finally, if the visual system uses only target template information to direct the first saccade, then in the picture cue condition the initial saccade should always be directed to the region containing the target object, regardless of scene context condition. In the word cue condition the visual system can only generate an abstract template, leaving initial saccades to be directed randomly.

The collective results would indicate whether both information types could be processed concurrently or if there were any bottle necks in processing cognitive information into eye movements, and also indicate when the visual system uses both target template and scene context information during a real-world search task. Though interpretation of the results would be limited to explaining only the initial stages of search, the results would provide a foundation from which more a complex series of search behaviours could be analysed.

A second research direction that could be taken with target template guidance is one of expected target position. Droll and Eckstein (2008) have shown that when a set of participants selected where in a scene they would expect to find a particular object, a second set of participants would tend to land near the mean expected position with the first fixation in a search task. The authors interpreted this set of results as evidence that the relative configuration of objects is rapidly extracted from a scene percept and used to direct gaze accordingly, and further evidence that the visual system uses expectation to guide eye movements.

This expected location research could be applied to target template guidance. Although not demonstrated as of yet, it would seem obvious that the expected target location in a scene varies with template specificity. For example, in Figure 15, 12 participants were asked to select where they would expect a cat to occur in the kitchen scene; half were given just the word cat (green spots) and half were given a picture of the cat in question (pink spots). As can be seen, the availability of a template shifts the mean expected position.

An interesting experiment would be to have participants choose expected locations as target specificity is manipulated and then run a second group of participants on a search task using the same cue manipulations and see if the initial fixation is directed toward the corresponding expected location.

If successful, the results would mean that we currently only half-understand target template guidance. Right now target template research interprets guidance in terms of searching for target-similar regions in a display. Success in the proposed experiment would suggest that a target template also affects an expectancy map: a map that suggests where in a scene the target might be. This map contributes then to the activation map which directs eye movements.

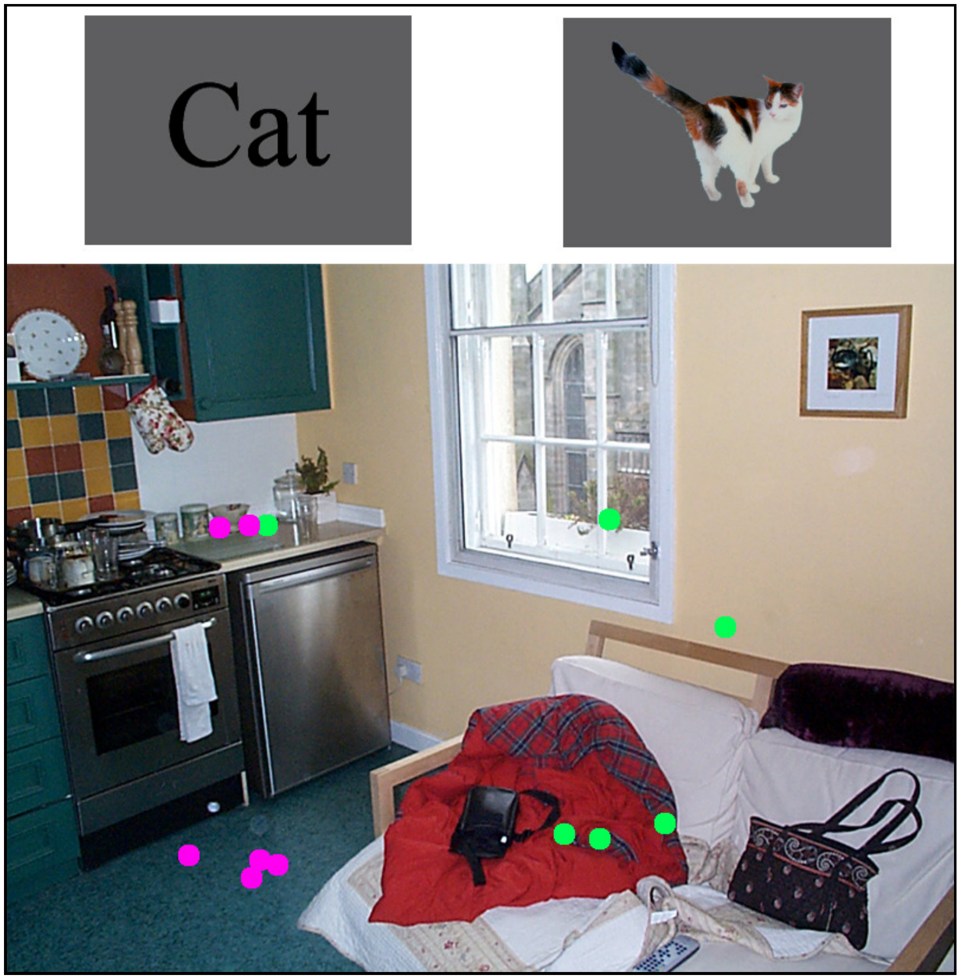

Figure 15 Twelve participants were asked to indicate where they thought the cat should be positioned in the scene. The spots in green correspond to the participants who were given just the word cat, while the spots in pink correspond to those who were given the picture of the cat.

The expectancy map would rely on an interplay between template and display percept, but instead of looking for similar regions between template and percept the expectancy map would be generated by the visual system looking for regions in a percept that fit within a cognitive knowledge structure that is affected by target knowledge. The more specific any of these variables, the more refined the hypothesised expectancy map becomes (see Equation 1 for a description). So as template specificity is increased the size of the expectancy map would decrease, resulting in a different saccade direction.

$$\text{Expectancy Map Size} = \frac{\text{(Display Size)}}{\text{(Display Size)(Cognitve Knowledge)(Target Template Specificty)(Task Certainty)}}$$

Equation 1 A descriptive equation outlining factors contributing to the hypothesised expectancy map. The more certain a participant is about the target's appearance, or the more information they have about the task or the scene's gist, the smaller the expectancy map will be. This changes where in the display a participant will direct saccades. Cognitive knowledge here refers to scene gist and layout.

If true, this particular process would be independent of the process of simply searching for features that match the target. This would mean that target template guidance works in two ways – selecting target similar regions and contributing to an expectancy map – and would necessarily redefine our understanding of a not just target template's role in real-world search, but the underlying mechanisms of the visual system. Of course, both rigorous testing and replication are needed before any effect of template specificity on expected target location can even begin to be assumed.

References

Antes, J. R. (1974). Time Course of Picture Viewing. *Journal of Experimental Psychology, 103*(1), 62-70.

Beutter, B. R., Eckstein, M. P., & Stone, L. S. (2003). Saccadic and perceptual performance in visual search tasks. I. Contrast detection and discrimination. *Journal of the Optical Society of America. A, Optics, image science, and vision, 20*(7), 1341-1355.

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene Perception: Detecting and Judging Objects Undergoing Relational Violations. *Cognitive Psychology, 14*, l43-l77.

Bravo, M. J., & Farid, H. (2009). The specificity of the search template. *Journal of Vision, 9*(1), 1-9.

Brockmole, J. R., Castelhano, M. S., & Henderson, J. M. (2006). Contextual Cueing in Naturalistic Scenes: Global and Local Contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*(4), 699-706.

Brockmole, J. R., & Henderson, J. M. (2006a). Recognition and attention guidance during contextual cueing in real-world scenes: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology, 59*(7), 1177-1187.

Brockmole, J. R., & Henderson, J. M. (2006b). Using real-world scenes as contextual cues for search. *Visual Cognition, 13*, 99-108.

Burgess, A. (1985). Visual signal detection. III. On Bayesian use of prior knowledge and cross correlation. *Journal of the Optical Society of America. A, Optics and image science., 2*(9), 1498-1507.

Buswell, G. T. (1935). *How People Look at Pictures*. Chicago: University of Chicago Press.

Carmi, R., & Itti, L. (2006). Visual causes versus correlates of attentional selection in dynamic scenes. *Vision Research, 46*, 4333–4345.

Castelhano, M. S., & Henderson, J. M. (2007). Initial scene representations facilitate eye movement guidance in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 753-763.

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing Task Influences on Eye Movements During Scene Perception. *Journal of Vision, 9*(3), 1-15.

Castelhano, M. S., Pollatsek, A., & Cave, K. R. (2008). Typicality aids search for an unspecified target, but only in identification and not in attentional guidance. *Psychonomic Bulletin & Review, 15*(4), 795-801.

Chen, X., & Zelinsky, G. J. (2006). Real-world visual search is dominated by top-down guidance. *Vision Research, 46*(24), 4118-4133.

Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology, 36*, 27-71.

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience, 18*, 193-222.

Droll, J., & Eckstein, M. P. (2008). Expected object position of two hundred fifty observers predicts first fixations of seventy seven separate observers during search. *Journal of Vision, 8*(6), 320a.

Eckstein, M. P., Beutter, B. R., Pham, B. T., Shimozaki, S. S., & Stone, L. S. (2007). Similar Neural Representations of the Target for Saccades and Perception during Search. *Neuron, 27*(6), 1266-1270.

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional Cues in Real Scenes, Saccadic Targeting, and Bayesian Priors. *Psychological Science, 17*(11), 973-980.

Egeth, H. E., Virzi, R. A., & Garbart, H. (1984). Searching for Conjunctively Defined Targets. *Journal of Experimental Psychology: Human Perception and Performance, 10*(1), 32-39.

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modeling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition, 17*(6/7), 945-978.

Einhäuser, W., Rutishauser, U., & Koch, C. (2008). Task-demands can immediately reverse the effects of sensory-driven saliency in complex visual stimuli. *Journal of Vision, 8*(2), 1-19.

Einhäuser, W., Schumann, F., Bardins, S., Bartl, K., Böning, G., Schneider, E., et al. (2007). Human eye-head co-ordination in natural exploration. *Network: Computations in Neural Systems, 18*(3), 267-297.

Findlay, J. M. (1997). Saccade Target Selection During Visual Search. *Vision Research, 37*(5), 617-631.

Fischer, M. H. (1999). An Investigation of Attention Allocation During Sequential Eye Movement Tasks. *The Quarterly Journal of Experimental Psychology, 52A*(3), 649-677.

Foulsham, T., & Underwood, G. (2007). How does the purpose of inspection influence the potency of visual saliency in scene perception? *Perception, 36*, 1123-1138.

Greenhouse, D. S., & Cohn, T. E. (1978). Effect of chromatic uncertainty on detectability of a visual stimulus. *Journal of the Optical Society of America, 68*(2), 266-267.

Hannus, A., van den Berg, R., Bekkering, H., Roerdink, J. B. T. M., & Cornelissen, F. W. (2006). Visual search near threshold: Some features are more equal than others. *Journal of Vision, 6*, 523-540.

Hayhoe, M. M. (2000). Vision Using Routines: A Functional Account of Vision. *Visual Cognition, 7*(1/2/3), 43-64.

Hayhoe, M. M., & Ballard, D. (2005). Eye Movements in Natural Behavior. *Trends in Cognitive Sciences, 9*(4), 188-194.

Henderson, J. M. (1992). Visual attention and eye movement control during reading and scene perception. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 260-283). New York: Springer-Verlag.

Henderson, J. M. (2003). Human gaze control in real-world scene perception. *Trends in Cognitive Psychology, 7*, 498-504.

Henderson, J. M. (2007). Regarding Scenes. *Current Directions in Psychological Science, 16*, 219-222.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual Saliency Does Not Account For Eye Movements During Visual Search in Real World Scenes. In R. P. G. van Gompel, M. H. Fischer, W. S. Murray & R. L. Hill (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 537-562). Oxford: Elsevier Ltd.

Henderson, J. M., & Ferreira, F. (1990). Effects of Foveal Processing Difficulty on the Perceptual Span in Reading: Implications for Attention and Eye Movement Control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 417-429.

Henderson, J. M., & Hollingworth, A. (1999b). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance versus visual salience during search for non-salient objects in real-world scenes. *Psychonomic Bulletin & Review, 16*(5), 850-856.

Henderson, J. M., & Pierce, G. L. (2008). Eye movements during scene viewing: Evidence for mixed control of fixation durations. *Psychonomic Bulletin & Review, 15*, 566-573.

Henderson, J. M., & Smith, T. J. (2009). How are eye fixation durations controlled during scene viewing? Further evidence from a scene onset delay paradigm. *Visual Cognition, 17*(6), 1055-1082.

Hollingworth, A. (2009). Two forms of scene memory guide visual search: Memory for scene context and memory for the binding of target object to scene location. *Visual Cognition, 17*(1/2), 273-291.

Intraub, H. (1981). Rapid Conceptual Identification of Sequentially Presented Pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7*(3), 604-610.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research, 40*(10-12), 1489-1506.

Itti, L., & Koch, C. (2001). Computational Modelling of Visual Attention. *Nature Reviews Neuroscience, 2*(3), 194-203.

Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254-1259.

Jiang, Y. V., & Wagner, L. C. (2004). What is learned in spatial contextual cueing: Configuration or individual locations? *Perception & Psychophysics, 66*, 454-463.

Joubert, O. R., Rousselet, G. A., Fize, D., & Fabre-Thorpe, M. (2007). Processing scene context: Fast categorization and object interference. *Vision Research, 47*, 3286–3297.

Judy, P. F., Kijewski, M. F., Fu, X., & Swensson, R. G. (1995). Observer detection efficiency with target size uncertainty. *SPIE, 2436*, 10-17.

Julesz, B. (1984). A brief outline of the texton theory of human vision. *Trends in Neuroscience, 7*(2), 41-45.

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition, 17*(6/7), 979 - 1003.

Keech, T. D., & Resca, L. (2010). Eye movement trajectories in active visual search: Contributions of attention, memory, and scene boundaries to pattern formation. *Attention, Perception & Psychophysics, 72*(1), 114-141.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology, 4*, 219-227.

Kunar, M. A., Flusberg, S. J., & Wolfe, J. M. (2006). Contextual cueing by global features. *Perception & Psychophysics, 68*(7), 1204-1216.

Land, M. F., & Hayhoe, M. M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research, 41*(25-26), 3559-3565.

Land, M. F., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception, 28*, 1311-1328.

Luria, S. M., & Strauss, M. S. (1975). Eye movements during search for coded and uncoded targets. *Perception and Psychophysics, 17*(3), 303-208.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11), 1-13.

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *Journal of Vision, 10*(2), 1-11.

Malcolm, G. L., Lanyon, L. J., Fugard, A. J. B., & Barton, J. J. S. (2008). Scan patterns during the processing of facial expression versus identity: An exploration of task-driven and stimulus-driven effects. *Journal of Vision, 8*(8), 1-9.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1995). Automatic control of saccadic eye movements made in visual inspection of briefly presented 2-D images. *Spatial Vision, 9*(3), 363-386.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1996). The relationship between the locations of spatial features and those of fixations made during visual examination of briefly presented images. *Spatial Vision, 10*(3), 165-188.

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation sequences made during visual examination of briefly presented 2D images. *Spatial Vision, 11*(2), 157-178.

Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805-1815.

Najemnik, J., & Geisler, W. S. (2005). Optimal eye movement strategies in visual search. *Nature, 434*, 387-391.

Najemnik, J., & Geisler, W. S. (2008). Eye movement statistics in humans are consistent with an optimal search strategy. *Journal of Vision, 8*(3), 1-14.

Navalpakkam, V., & Itti, L. (2005). Modeling the influence of task on attention. *Vision Research, 45*(2), 205-231.

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research, 46*(5), 614-621.

Oliva, A., & Schyns, P. (2000). Diagnostic colors mediate scene recognition. *Cognitive Psychology, 41*, 176–210.

Oliva, A., Wolfe, J. M., & Arsenio, H. C. (2004). Panoramic Search: The Integration of Memory and Vision in Search Through a Familiar Scene. *Journal of Experimental Psychology: Human Perception and Performance, 30*(6), 1132-1146.

Palmer, S. E. (1999). *Vision science: photons to phenomenology*. London: MIT Press.

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision Research, 42*(1), 107-123.

Potter, M. C. (1975). Meaning in visual scenes. *Science, 187*, 965–966.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology, 81*, 10-15.

Rajashekar, U., Bovik, A. C., & Cormack, L. K. (2006). Visual search in noise: Revealing the influence of structural cues by gaze-contingent classification image analysis. *Journal of Vision, 6*, 379-386.

Rao, R. P. N., Zelinsky, G. J., Hayhoe, M. M., & Ballard, D. H. (2002). Eye Movements in Iconic Visual Search. *Vision Research, 42*, 1447-1463.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Rizzolatti, G., Riggio, L., Dascola, I., & Umiltá, C. (1987). Reorienting attention across the horizontal and vertical meridians: evidence in favor of a premotor theory of attention. *Neuropsychologia, 25*(1A), 31-40.

Rothkopf, C. A., Ballard, D. H., & Hayhoe, M. M. (2007). Task and context determine where you look. *Journal of Vision, 7*(14), 1-20.

Schmidt, J., & Zelinsky, G. J. (2009). Search guidance is proportional to the categorical specificity of a target cue. *Quarterly Journal of Experimental Psychology, 62*, 1904-1914.

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 14*(7), 1-17.

Tatler, B. W., Baddeley, R. J., & Gilchrist, I. D. (2005). Visual correlates of fixation selection: effects of scale and time. *Vision Research, 45*, 643-659.

Tatler, B. W., & Vincent, B. T. (2008). Systematic tendencies in scene viewing. *Journal of Eye Movement Research, 2*(2), 1-18.

Tatler, B. W., & Vincent, B. T. (2009). The prominence of behavioural biases in eye guidance. *Visual Cognition, 17*(6-7), 1029-1054.

Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2007). An efficient technique for revealing visual search strategies with classification images. *Perception & Psychophysics, 69*(1), 103-112.

Tavassoli, A., van der Linde, I., Bovik, A. C., & Cormack, L. K. (2009). Eye movements selective for spatial frequency and orientation during active visual search. *Vision Research, 49*, 173-181.

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

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

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

Tseng, P.-H., Carmi, R., Cameron, I. G. M., Munoz, D. P., & Itti, L. (2009). Quantifying center bias of observers in free viewing of dynamic natural scenes. *Journal of Vision, 9*(7), 1-16.

Turano, K. A., Geruschat, D. R., & Baker, F. H. (2003). Oculomotor strategies for the direction of gaze tested with a real-world activity. *Vision Research, 43*, 333-346.

Unema, P. J. A., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12*(3), 473-494.

van Diepen, P. M. J., Wampers, M., & d'Ydewalle, G. (1998). Functional division of the visual field: Moving masks and moving windows. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 337-355). Oxford: Elsevier.

Velichkovsky, B. M., Joos, M., Helmert, J. R., & Pannasch, S. (2005). *Two visual systems and their eye movements: evidence from static and dynamic scene perception.* Paper presented at the Cognitive Science Society, Stresa, Italy.

Vickery, T. J., King, L. W., & Jiang, Y. H. (2005). Setting up the target template in visual search. *Journal of Vision, 5*(1), 81-92.

Williams, D. E., & Reingold, E. M. (2001). Preattentive guidance of eye movements during triple conjunction search tasks: The effects of feature discriminability and saccadic amplitude. *Psychonomic Bulletin & Review, 8*(3), 476-488.

Williams, L. G. (1967). The effects of target specification on objects fixated during visual search. *Acta Psychologica, 27*, 355-360.

Wolfe, J. M. (1994a). Guided Search 2.0: A revised model of visual search. *Psychonomic Bulletin & Review, 1*(2), 202-238.

Wolfe, J. M. (1994b). Visual Search in Continuous, Naturalistic Stimuli. *Vision Research, 34*(9), 1187-1195.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science, 9*(1), 33-39.

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.

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience, 5*, 1-7.

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research, 44*, 1411-1426.

Wolfe, J. M., O'Neill, P., & Bennett, S. C. (1998). Why are there Eccentricity Effects in Visual Search? *Perception & Psychophysics, 60*(1), 140-156.

Wolfe, J. M., Oliva, A., Horowitz, T. S., Butcher, S. J., & Bompas, A. (2002). Segmentation of objects from background in visual search tasks. *Vision Research, 42*, 2985-3004.

Wolfe, J. M., Reijnen, E., Van Wert, M. J., & Kuzmova, Y. (2009). In visual search, guidance by surface type is different than classic guidance. *Vision Research, 49*, 765-773.

Yarbus, A. (1967). *Eye Movements and Vision*. New York: Plenum Press.

Zelinsky, G. J. (1996). Using Eye Saccade to Select the Activity of Search Movements. *Vision Research, 36*(14), 2177-2187.

Zelinsky, G. J. (2008). A Theory of Eye Movements During Target Acquisition. *Psychological Review, 115*(4), 787-835.

Zelinsky, G. J., Rao, R. P. N., Hayhoe, M. M., & Ballard, D. H. (1997). Eye Movements Reveal the Spatiotemporal Dynamics of Visual Search. *Psychological Science, 8*(6), 448-453.

Zelinsky, G. J., Zhang, W., Yu, B., Chen, X., & Samaras, D. (2006). The role of top-down and bottom-up processes in guiding eye movements during visual search. In Y. Weiss, B. Scholkopf & J. Platt (Eds.), *Advances in Neural Information Processing Systems* (Vol. 18, pp. 1569-1576). Cambridge, MA: MIT Press.

Appendix

The contents of Chapter 2 were published in *Journal of Vision* in 2009, while the contents of Chapter 4 were published in *Journal of Vision* in 2010.

Malcolm GL, Henderson JM. "The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements." *Journal of Vision*, 2009, 9(11):8, 1-13.
http://journalofvision.org/9/11/8/

Malcolm GL, Henderson JM. "Combining top-down processes to guide eye movements during real-world scene search." *Journal of Vision*, 2010, 10(2):4, 1–11.
http://journalofvision.org/10/2/4/