University of Dundee

DOCTOR OF PHILOSOPHY

Familiarity in natural behaviour

Effect of task, objects and environment on gaze allocation

Scrafton, Sharon

Award date:
2015

Familiarity in natural behaviour: Effect of task, objects and environment on gaze allocation

Sharon Scrafton, MA

Thesis submitted to the University of Dundee for the
Degree of Doctor of Philosophy

December, 2015

Table of Contents

Chapter 5 Comparing eye movements during tea making in both novel and familiar natural environments 119

Table of Figures

Acknowledgements

To my supervisor, Professor Ben Tatler – you have essentially been the vestibulo-ocular movements of my thesis – keeping me steady and helping me focus. Had it not been for your enthusiasm for vision science, your encouragement, your insight and patience, my thesis (and in fact life) would have taken a completely different direction. You and this subject have won me over, from my undergraduate disappointment of being assigned a vision related dissertation topic and supervisor, to becoming one of you! You have patiently put up with my insistence of completing all of my experiments using the mobile eye tracker and then resisted the urge to say “I told you so” when it came to the many months of manual coding. More than that, you have encouraged me every step of the way and believed in me, your support is so greatly appreciated. Thank you.

My thanks to the School of Psychology in Dundee for the studentship allowing me to pursue my studies. In particular, I would like to thank Astrid Schloerscheidt for the professional and personal support. To John Morris, the head technician during my time at Dundee – you saved all of my raw and coded data from the belly of a broken hard-drive, without you there would literally be no thesis. To the participants who took part in my studies, for not much in return, they willingly gave up their time, wore strange glasses, and poured and brewed many cups of tea. I will raise a cup of tea in their honor.

To Matt, I think we missed a trick when our vows talked about *in sickness and in health* and *for richer for poorer* – what they should have said was *through your thesis and through mine,* nevertheless we made it!!!!! For everything, my eternal thanks and gratitude.

My daughter Delilah – at the age of 3 years you inspire me go on to try to be the best scientist I can, your insightful questions, your tenacity, your thirst for knowledge and your delight in sharing your learning is what our job is all about. Thank you.

Declaration

This thesis is a presentation of my original research work. Wherever contributions of others are involved, every effort is made to indicate this clearly. All references cited have been consulted with and given due reference.

The work completed for this thesis is all my own work and has not previously been accepted for a higher degree.

Sharon Scrafton

Abstract

For active tasks we have to appropriately allocate our gaze spatially and temporally so that we are fixating informative areas when the crucial information is available. We know that vision supports action, and several fundamental elements of how this is so have been established, for example that the eye leads the hand during action. What we do not know is whether the spatiotemporal allocation of gaze is consistent regardless of the task, the objects and the level of familiarity we have with the environment.

In Chapter 3 we found that visual behaviour changes as a result of the task being undertaken, with more looks to task irrelevant objects, longer eye-hand latencies and more visually guided putdowns of objects made for tea making rather than sandwich making. Analysis revealed that the objects used in the two different tasks did affect eye-hand latencies. In Chapter 4 this issue was explored further and it was found that the properties of objects (glasses) such as glass type (where height may be the important factor) influenced visual guidance if the glass was empty during the set down, but that level of liquid contained, and the material it was made from impacted the likelihood of using visual guidance for a second putdown of the same glass. These results indicate flexibility in terms of the allocation of visual guidance depending on our knowledge of the object properties, and suggest that risk may be an important factor in this.

The effect of familiarity with an environment was looked at in three ways. First in Chapter 5 we compared people making tea in familiar environments (their own kitchens) and in novel environments (their experimental partners kitchen).

Second we explored the acquisition of familiarity by having participants perform a task in the same environment for 10 consecutive days (Chapter 6) and finally we investigated what information was encoded incidentally by having the participants from Chapter 6 perform a new task in the same environment for two subsequent days. We found that people were faster to complete the same task in a familiar environment than a novel one but that it was not just that search was facilitated and thus shorter, visual behaviours such as visual exploration and looks to task irrelevant objects were fewer when in familiar environments and several elements of the Object Related Action (ORA) also reduced in a temporal nature. We found that during the acquisition of familiarity people encoded information about the layout of objects in the scene which facilitated search in Chapter 7 but there appeared to be no such effect on the ORA, suggesting that object specific information for task irrelevant objects is not incidentally encoded.

The findings of this thesis suggest that spatiotemporal allocation of gaze in natural tasks depends on the context of the environment, the properties of objects and our level of prior knowledge.

Chapter 1 General introduction

1.1 Overview

If the goal of psychology is to understand the mind and behaviour, then the study of our visual behaviour is crucial since so much of the way we act is influenced by what we see or think we see. We are active agents interacting with the world around us and the common thread in all of our behaviours is that they all involve movement. Eye movements, communication (spoken, gestural and written), and manipulations of objects and our environment are the typical measures that psychologists take as indicators of cognitive output. According to Wolpert, Ghahramani, and Flanagan (2001), the entire purpose of the human brain is to produce movement. All of our interactions with the world are ultimately governed by the motor system and behaviour is inherently made up of movement, with our visual behaviour being no different: we move our eyes in order to perceive the world around us and enable us to act upon it (Land & Tatler, 2009). The study of eye movements allows us to examine the input (perception of the world around us), the cognitive processes involved in dealing with our perceptions, and the impact on our behaviours (the output). Indeed the limited spatial and temporal sampling of the eyes means that they impart a perceptual bottleneck on the information that can be gathered and conveyed to the brain. Where we fixate is under our voluntary control and we make around 130,000 fixations in a 12-hour period of wakefulness (based on the assumption of 3 fixations per second). Furthermore, vision is restricted in space and time and thus must be allocated precisely and strategically to get the information we need at the time that we need it. Therefore, the factors that drive these fixations, which then impact the way we behave in the world, are of the upmost interest to psychologists. Our eye movements in the context of our interactions with objects and environments during natural active tasks are the focus this thesis.

Vision is an active system, constantly gathering information from the environment around us. We move our eyes to direct the foveae to areas of interest to acquire information required to support interactions with our environment (Findlay & Gilchrist, 2003; Land and Tatler, 2009). Vision is not a passive sense merely receiving information presented to it, rather the eyes produce a complex dance of movements in order to support and guide our behaviour in the world. We know from early work (Buswell, 1935; Yarbus, 1967) and from more modern studies (Land et al, 1999; Hayhoe et al, 2001, 2003) that observers direct their gaze towards task relevant areas of scenes or environments. Typically, in everyday life, many of the actions we perform are fairly complex and sequential with an overriding goal made up of smaller subtasks; studies examining eye movements in the context of these natural behaviours have revealed the complex role our vision plays in supporting these actions. In fact, we know that vision typically leads action, by about a second, (Buswell, 1920 Weaver, 1943, Land et al., 1999) and is tightly coupled with whatever action is occurring at any given point in a task. Land, Mennie and Rusted (1999) defined the fixations made during a task by their purpose; *locating, directing, checking and guiding,* finding that they occupy rather a large amount of our visual behaviour for even the apparently automated task of tea making. Thus we can conclude that eye movements during a natural task are made by an active system, seeking out the information needed for each element of the task (Findlay & Gilchrist, 2003; Land & Tatler, 2009).

Fixations are made on a voluntary basis, the region we saccade to and then the object we fixate is under our control. Therefore, the locations that we select with the visual system can give us some insight into the information we are extracting from the world to complete a task. In fact, Buswell (1935) and Yarbus (1967) showed that when an observer was asked to view a scene with a set of instructions in mind, the patterns of eye movements varied considerably depending on the instructions given. For example, Yarbus demonstrated when the participant was asked to make a judgment about the age of a figure in the painting, the pattern of eye movements differed substantially from when they were asked to remember features about the clothes worn. These are both examples of the earliest pieces of work supporting the notion that we purposefully use our active vision to seek out information rather than merely reacting to visual stimuli.

Although *task* has perhaps been one of the most extensively studied issues influencing visual strategies for information gathering, other factors play important roles in guiding the eyes. Learned knowledge of physical laws that govern scene organization, such as gravity and support (Biederman, Mezzanotte, & Rabinowitz, 1982), play an important role in scene perception and inspection, and we are sensitive to unrealistic violations of these laws (Biederman, 1976; Biederman et al, 1982; Vo & Wolfe, 2013; Võ & Schneider, 2010; see Figure 1.1). Similarly, we look where we expect to find objects in scenes based upon past experience (Torralba et al., 2006). So for example, we may have a set of expectations (or schema, Bartlett, 1932) tied to the environment of ‘kitchens’ in general, in that we would likely have a rough idea of the typical layout and the objects ordinarily contained in a kitchen. These expectations are built on our experiences with the world, and would likely guide our visual behaviour so much so that we would look in different places in a kitchen than we would for example in a field (Shallice, 1988).

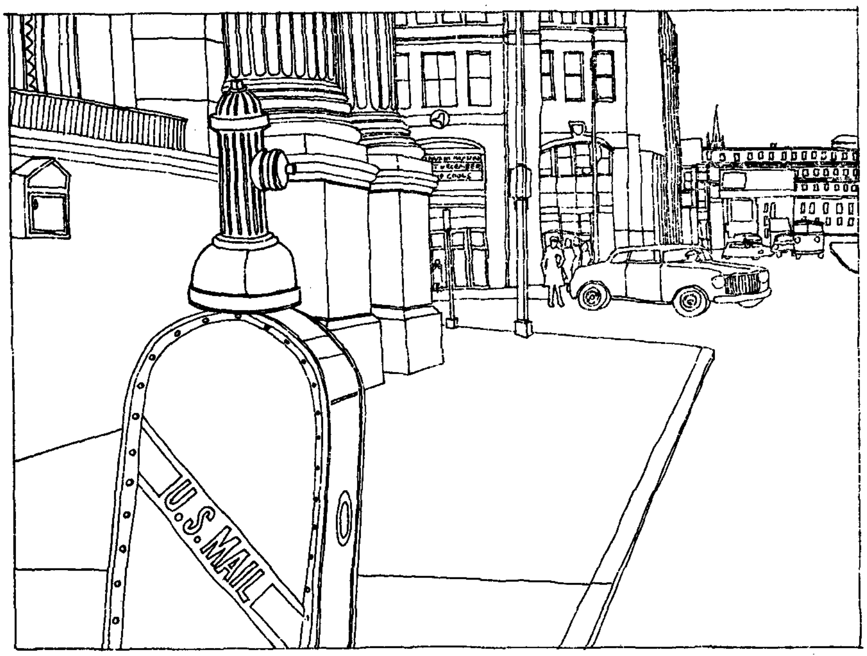


Figure 1.1. Unrealistic violations of expectation in a scene from Biederman (1976).

Prior exposure to a scene has in fact been shown to impact search behaviour. Brockmole and Henderson (2006) repeatedly showed photographs of scenes that contained objects that were consistently but arbitrarily placed, the images were then flipped to a mirror image. The results showed that the observers would initially look at the original (expected) location for the target object but then quickly move to look at the new location, which the authors noted made search slow a little however, overall savings were made. Similarly Võ & Wolfe (2012) demonstrated that if a subject searched for an object in a scene, the subsequent search for the same object was speeded dramatically despite many intervening searches. These studies demonstrate that experience with a scene can impact our visual behaviour, even to the extent that we rely on our expectations to make anticipatory eye movements to where we think an object will be rather than use the visual information available in the very first instance (Brockmole & Henderson, 2006).

Our brain takes advantage of common associations among objects in the environment to facilitate visual perception and cognition (Bar, 2004). Having expectations about which objects typically occur in which contexts have been referred to as schemata (Bartlett, 1932; Biederman, 1974; Piaget, 1955), contextual effects (Palmer, 1975), scripts (Schank, 1975) and frames (Minsky, 1975). The types of contextual expectations that we might have about a scene were characterized by Biederman, Mezzanotte, and Rabinowitz (1982) as including ‘support’ (most objects are physically supported rather than float), ‘interposition’ (for example, occlusion), ‘probability’ (the likelihood that certain objects will be present in a scene), ‘position’ (the typical positions of some objects in some scenes) and ‘size’ (the familiar relative size of objects). Since these expectations are built on experience one might think that it would take a long time to acquire the experience with which expectations can be set, however work looking at expectations and in particular violations of expectation in infants reveals that infants as young as four months have expectations about most of the characteristics of objects identified by Biederman et al., (1982); support (Hespos & Baillargeon, 2008), occlusion (Aguiar & Baillargeon, 2002), position, (Wang, Baillargeon, & Paterson, 2005), size (Wilcox, 1999).

The expectations that we set up for a scene and the objects and properties of those objects in the scene are so influential that we actually have difficulty coming back from violations of these expectations, in fact they are processed more slowly and less accurately (Biederman et al., 1982). Not only do we find it difficult to recover from expectation violations, contextual cuing can be so influential that we are even susceptible to confabulating the presence of objects if we had been primed to consider a certain environment that we have experience with. Brewer and Treyens (1981) demonstrated that by referring to a waiting room as an office, not only were individuals more likely to report having seen office related objects that were present in the room but they reported having seen objects that one would expect to have found in an office that were not actually present.

Contextual expectations clearly influence why we would look in a certain place when searching for an object, they also affect our perception and memories of scenes and objects. We appear to formulate many of these expectations from a very young age and clearly benefit on a processing level from utilising them, however, the literature so far has tended to rely on exploiting generic schematic expectations, the focus of this thesis is to examine the effect of level of prior knowledge on our eye movements and visual behaviour.

1.2 How do we move our eyes?

Our experience of the way we visually perceive the world is one of a continuous, fluid, rich perception. We are unaware that our region of clear vision is limited to approximately the size of a thumbnail held at arm's length

(Land, 1999). Clarity is presumed from our viewing experience, however in order to compensate for the relatively small area of around ~2° known as the fovea (Steinman, 2003; see Figure 1.2) we have to move our eyes up to three times per second in order to gather accurate visual information. Due to the low acuity of our peripheral vision, we use eye movements to direct our fovea to the location in the world where we wish to acquire information from. The moments of stability that separate eye movements are referred to as fixations and last for roughly 300 ms (Rayner, 1998; Land, 1999). The movements that we make from one area to another are rapid rotations known as saccades, a term often credited to being first used by Javal (1879). Saccades are ballistic eye movements which can reach velocities of $1000°s^{-1}$ (Carpenter, 1988). One of the most fascinating accidental findings regarding saccades was made by Erdmann and Dodge (1898) who noted that whilst watching eye movements using a mirror, during a reading task, they were unable to see their own eye during a saccade, concluding that we are effectively blind during a saccade. This finding was extremely important, since up until that point, it was thought that perception continued throughout the movements of the eyes (Cattell, 1900). Not only did this finding regarding saccades suggest this was not the case, Dodge (1900) also demonstrated the importance of keeping the eye still in between saccades, since in fact it is when the eye is stationary that we extract visual information

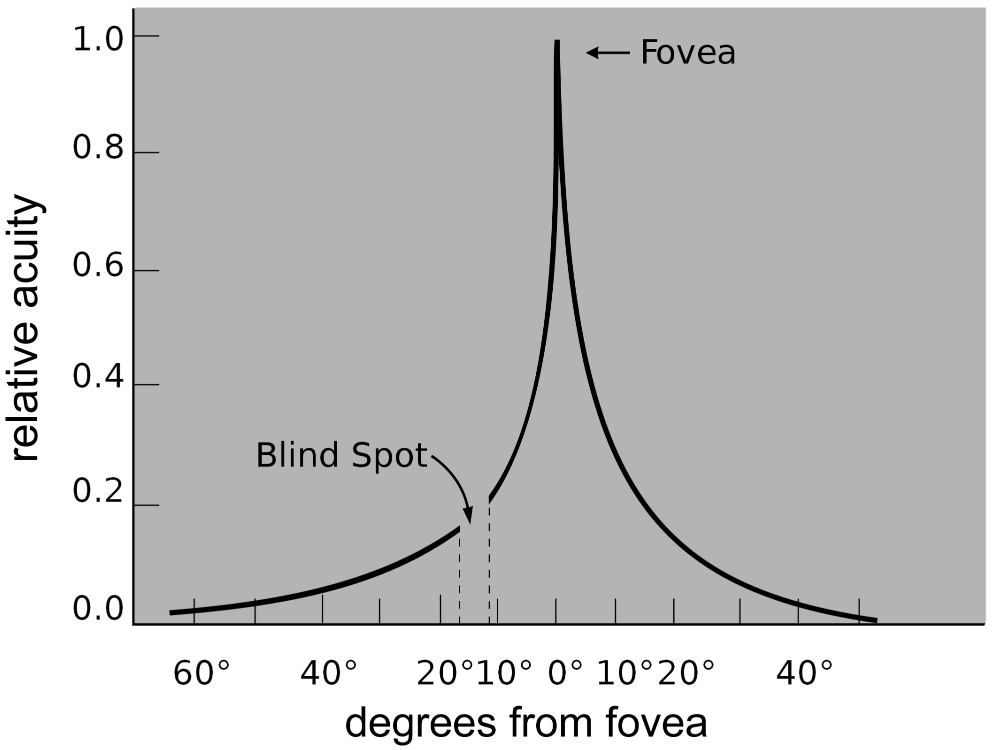


Figure 1.2. Relative acuity across the human retina, adapted from Hunziker (2006).

Specifically, when we want to extract information from an area we are not currently foveating, we are required to make an eye movement to that area. When gaze is fixated on the target, we must employ movements which hold steady our gaze such as the vestibulo-ocular reflex and optokinetic reflex (Land & Tatler, 2009) which compensate for head and body movements in order to extract the necessary information without blur, since the process of photoreception is slow, taking approximately 20 ms for a cone to respond fully to a step change in the light reaching it (Findlay & Gilchrist, 2003; Friedburg, Allen, Mason, & Lamb, 2004; Land & Tatler, 2009). In addition to these are eye movements whose purpose is to correct either position of the eyes and maintain the relationship between both eyes to resolve disparity (vergence eye movements) or to compensate for changes in head and body position (i.e. the vestibular-ocular reflex). There will be some circumstances where we may be inclined to track a moving object. In this case we move our eyes *with* the object, known as pursuit eye movements. Typically in visual behaviour, periods of fixation on objects are punctuated with saccadic relocations. The gaze stabilizing movements that we employ essentially support our 'fixate and saccade' strategy which is according to Land (1999) the main way we view the world.

In real environments, we encounter situations where there is a need to relocate gaze to a location far from the current target of our central vision (e.g., Land, Mennie, & Rusted, 1999; Tatler & Land, 2011). In such situations it is not possible to orient to a region using movements of our eyes alone, since the oculomotor range only extends to about +/-55° (Guitton & Volle, 1987). Outside of this, head and possibly body movements (such as trunk rotations) must be implemented (Land, 2004). Early studies, during 1950s typically focused on the eye relocations only oftentimes with the head of the viewer being artificially supported and held stationary, which is in fact still the case in many studies particularly those examining eye movements when reading (Kowler, 1990). However, apart from when we are reading a book or working at a computer (even then of course we still move our heads), it is normal for humans to make both head and body movements frequently in order to visually orient to areas on interest, therefore studying eye movements in isolation of head and body movements does not complete the picture of our visual behaviour.

1.3 Why do we move our eyes?

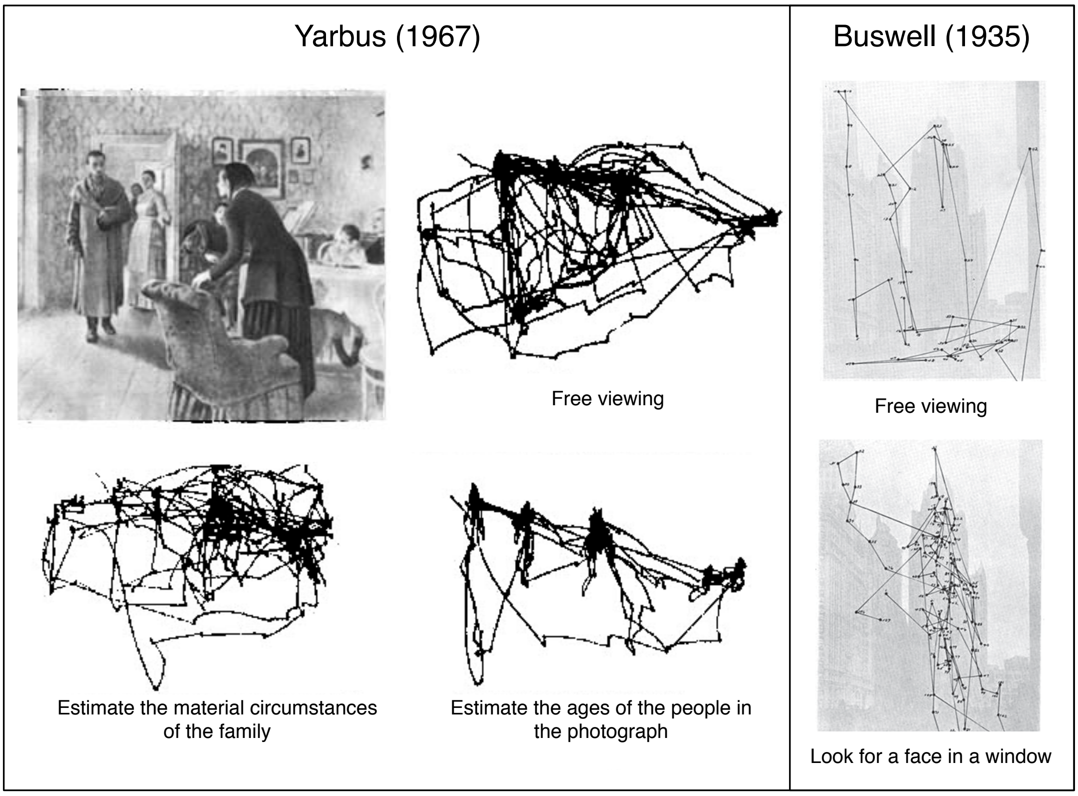
Aside from the physical limitations that require us to move our eyes in order to see clearly, the factors that drive our eyes to fixate different objects and areas in scenes and environments have been considered in depth in the literature over the years. The arguments can be split roughly into two camps, those that propose we move our eyes in response to external factors (something present in a scene for example) or those that argue we are motivated to move our eyes by internal factors (for example our desire to find information to support a goal). As early as 1935, Buswell noted that certain areas of images were consistently fixated more than others and referred to these areas as 'centers of interest', which he proposed had two possible explanations as to what made these areas interesting, one being that the stimulus contained something that attracted the eye or the second option that the viewer was motivated by some sort of cognitive 'interest' in certain parts of the scene. Although there may be instances where something in a scene or environment is inherently interesting, typically what is ‘interesting’ to us at any given time is information that we need to know for a specific purpose. Buswell and later Yarbus (1967) demonstrated that in fact it was possible to change the place a viewer looked at by giving different viewing instructions, thus supporting Buswell's second reasoning of the motivation for looking at certain areas over others - we are driven by top down factors and have our own cognitive motivations for choosing to fixate some areas over others depending on the information available there.

Tatler (2009) points out that it is surprising that this very early work clearly demonstrating the importance of top down factors in in our visual behaviour when viewing complex scenes was then somewhat overshadowed by work proposing that the visual system is driven largely by bottom up factors. However, rather than the low-level features in images as proposed by the saliency model (Itti & Koch, 2000), salient stimuli can be considered to refer to surprising events (Knudsen, 2007). These types of salient events are considered by Knudsen to be rare and essentially distracting from the task in hand, like for example a sudden flash of light. It would seem plausible that events like this would draw our gaze and visually attending to these unexpected events would reliably occur. However, several studies of explicit attentional capture reveal a surprising degree of blindness to salient or unusual events that we might expect to capture attention. For example, observers often fail to notice surprisingly large, but unexpected changes to their visual world, such as a change to the identity of the central actor in a brief motion picture (Levin & Simons, 1997; Simons, 2000) Furthermore observers sometimes fail to notice an unexpected object or event altogether – a phenomenon now known as ‘inattentional blindness’ (Mack, Rock, & Press, 1999; Mack, Tang, Regina, & Kahn, 1992; Newby & Rock, 1998; Simons & Chabris, 1999). Ultimately then one could argue that even something as powerful as an unexpected event or change in object is not enough to explain why we look at something, since unless a viewer is expecting an event (an internal cognitive state) we can essentially be blind to change.

Other factors in scenes can explain fixation selections more accurately than the salience model, for example object-level information (Einhauser et al., 2008) and in fact even the predisposition of viewers in the way that they move their eyes in scene viewing (Tatler & Vincent, 2009). Although many of these studies, including the traditional saliency model, are referring to a more spatial account of the allocation of gaze, the reason of *why* we look where we do is tightly bound with the question of *where* we look. The importance of task, expectations and information extraction in shaping our motivation to move our eyes and seek out information points to a goal driven visual system, thus we can conclude that it is unlikely to be simply down to low level scene features alone. Although some threads of current research maintain the importance of low-level features driving gaze selection (for example, Ehinger, Hidalgo-Sotelo, Torralba, & Oliva, 2009; Judd, Durand, & Torralba, 2012; Kanan, Tong, Zhang, & Cottrell, 2009; Risko, Laidlaw, Freeth, Foulsham, & Kingstone, 2012; Torralba, Oliva, Castelhano, & Henderson, 2006) it is becoming increasingly accepted that salience only very minimally explains our guidance of gaze, that in fact our gaze control is influenced considerably more by top-down factors and that the reasons why we move our eyes are motivated internally by high level cognitive factors.

1.4 Where do we move our eyes

We know from very early eye movement research by Buswell (1935) and Yarbus (1967) that the task we are completing has a strong influence on where in a scene we choose to fixate. Both of these early pieces of work clearly demonstrated that when viewers had a particular question in mind, the regions of fixation were tightly coupled to the areas in the scene where the required information was located. Buswell (1935) compared eye movements of participants viewing a scene of the Tribune tower in Chicago, first they were asked to free view the photograph and then to look at it and locate a person looking out of the window in the scene. The scan paths produced for each condition look quite different (see Figure 1.3). In the free viewing condition, the scan paths were fairly spread around the image with few fixations on the actual tower, however when participants viewed the image in the second condition, the fixations were significantly more concentrated on the tower and in particular on the windows in the tower. These findings were extended by Yarbus who had the subject view an image with several different questions in mind, finding that each question yielded a different set of scan paths tightly related to the relevant areas in the scene.



Yarbus (1967)

Buswell (1935)

Free viewing

Free viewing

Estimate the material circumstances of the family

Estimate the ages of the people in the photograph

Look for a face in a window

Figure 1.3. Eye movements from Yarbus (1967, left panel) and Buswell (1935, right panel) showing eye movement differences with different tasks.

Where we look is not only influenced by the fact that we are engaged in a task but also the type of task we are actually carrying out. Castelhano, Mack, and Henderson (2009) demonstrated that viewers' fixations differed depending on whether they were asked to view a scene on which they would then be tested about the specific objects, compared with viewing a scene to search for an item. In the memorization task, viewers tended to spread fixations across all of the objects in a fairly even manner, whereas in the visual search condition fixations tended to be concentrated on areas which would most likely contain the target object, consistent with evidence showing that context information, (for example looking for clocks on walls as it is the most likely location to find one), leads to more efficient searches (Brockmole, Castlehano & Henderson, 2006; Castlehano & Henderson, 2007; Neider & Zelinsky, 2006).

Research has suggested that guidance of the eye based on understanding of scene gist might occur from the very first saccade after an image is presented, although some debate exists about what we can tell from the first fixation. Initially Mackworth & Morandi (1967) and Antes (1974) proposed that semantic informativeness drives the first fixation in scene viewing, which suggests that participants are very quickly able to process a scene's characteristics and direct fixations to information rich areas very quickly. At first glance this finding is not entirely surprising since we know that viewers can extract gist in an incredibly short amount of visual inspection time (Potter, 1975). However, several studies since then (De Greaf et al., 1991; Henderson et al., 1999) have not replicated this effect of initial guidance based on semantic informativeness and instead argue that visual informativeness, i.e. low-level features may be the driving factor in these first fixations (e.g., Henderson, & Hollingworth, 1999). Although semantic informativeness does not seem to drive our first fixation, subsequent fixations do indeed seem to be concentrated on semantically informative regions (Loftus & Mackworth, 1978; Henderson et al., 1999). Furthermore we seem to use our prior knowledge throughout a viewing task in order to guide and constrain where regions of interest in a scene are likely to be. For example, Neider and Zelinsky (2006) asked participants to search for objects in pseudo-realistic scenes to examine the effect of context. The target objects used in the experimental scenes would typically be constrained to certain areas in real life (for example a hot air balloon in the sky, or a car on the ground), thus viewers would likely have expectations about where to search based on this prior knowledge. The results demonstrated that participants did indeed constrain their search to areas in the scene where the object would be likely to be and that target present in the expected area induced faster search times. However when the target object was not present, search was less restricted to the expected areas, demonstrating that not only do we use context to direct where to look but also the system is flexible enough to cope with expectation violations.

Expectations appear to guide our fixations to areas in the scene that are likely to hold the objects we are interested in. Typically the objects that are interesting are the ones related to the task in hand. One body of work has examined whether we look at objects that are semantically inconsistent. The findings are mixed; several authors have found that we do indeed make earlier fixations to incongruent objects (Friedman, 1979; Loftus & Mackworth, 1978; Becker, Pashler & Lubin, 2008), whereas some recent pieces of work have failed to replicate these findings (Brockmole & Henderson, 2008; De Graef, 1998; De Graef et at, 1990; Henderson, Weeks & Hollingworth, 1999).

Spotorno, Malcolm, and Tatler (2014), manipulated knowledge during a visual search task, both about the target itself and the target's location during search of a real-world scene, in order to examine the effect on the initial period of viewing and across the subsequent phases of scene viewing. The authors found a higher concentration of fixations were made to the target object following cuing with a picture of the actual target and when the target was in the expected 'normal' arrangement. The authors argue that if we have access to detailed information about the features of a search target, we can use this to find objects effectively even when they are not in the expected location. In a subsequent paper, the authors (Spotorno, Malcolm, & Tatler, 2015) focused on the effect of misleading expectations on visual searches. The author's findings suggest that the visual system can flexibly adopt an oculomotor strategy that utilises multiple sources of high-level guidance during a search task.

Whilst it may be tempting, considering the evidence presented above, to conclude that only high-level cognitive factors related to task influence where people look when viewing a scene, we also have to bear in mind that we as viewers have biases in terms of where we put our gaze. Not only do observers have a tendency to fixate near the centre of the screen (Tatler, 2007) they also tend to concentrate looks to foreground objects (Vincent, Baddeley, Correani, Troscianko, & Leonards, 2009). However that is not to say that viewing tendencies are not high-level, in fact one of the speculated reasons as to why we look in the centre of the screen and at foreground objects may be that we know that typically the most informative region of the screen is the centre and that foreground objects tend to be the most important ones in a scene.

Having understanding of the factors that drive our gaze around a scene allows the development of models that take into account these high-level factors that have been shown to influence where we look. However not all models are based on a purely top-down approach, models arguing a neuropsychological standpoint, bottom-up and modulated bottom-up models continue to emerge. One neuropsychological approach, taken by Findlay and Walker (1999) was to look at the neural pathways involved in the control of visual behaviour. They proposed a model which was based on pathways in the brain and comprised a fixate centre and a move centre, linked by lateral inhibition. The authors argued that when there is activation in the fixation centre, we inhibit saccades; conversely when we activate the movement centre, we are able to disengage from the fixation and program a new saccade. During the fixation, a saliency map develops that represents the spatial locations of potential saccade targets that determine where we look.

The saliency map model proposed by Itti and Koch (2000; derived from Koch & Ullman, 1985) is a bottom up model of visual attention which involves the computational application of a saliency map mechanism that guides attention to select the most ‘salient’ regions (based on purely low-level differences in colour/orientation/intensity) of an image or scene (see Figure 1.4 for an illustration of the organisation of the saliency map). Despite this model gaining significant popularity for a number of years, mounting evidence over the last decade or so demonstrates that the salience model does not explain our visual behaviour adequately and in fact accounts for only a modest proportion of the variance in our fixations when viewing scenes (Tatler, Land, Hayhoe & Ballard, 2011; Hayhoe, 2007; Tatler 2009; Land & Tatler, 2009, Schutz, Braun &

Gegenfurtner, 2011). The salience model may go some way to explain why we would allocate our gaze to areas in a scene when we are viewing static arrays with non-complex visual features and in fact it has been found to predicting gaze slightly better than chance for complex natural scenes (Foulsham & Underwood, 2008). However this is based on correlations between low-level image properties and the fixations made to them and as has been subsequently pointed out, these correlations do not imply that the features of parts of the scene cause the allocation of fixations (Henderson, 2003; Henderson, Brockmole, Castlehano & Mack, 2007; Tatler, 2007). Furthermore it has been demonstrated that when the viewer's task is manipulated, predictive powers of the model vanish (Foulsham & Underwood, 2008; Henderson et al., 2007). Subsequent models have typically included saliency as one factor that may contribute to visual guidance and in fact several models have included the salience map and modified the model to include other factors.

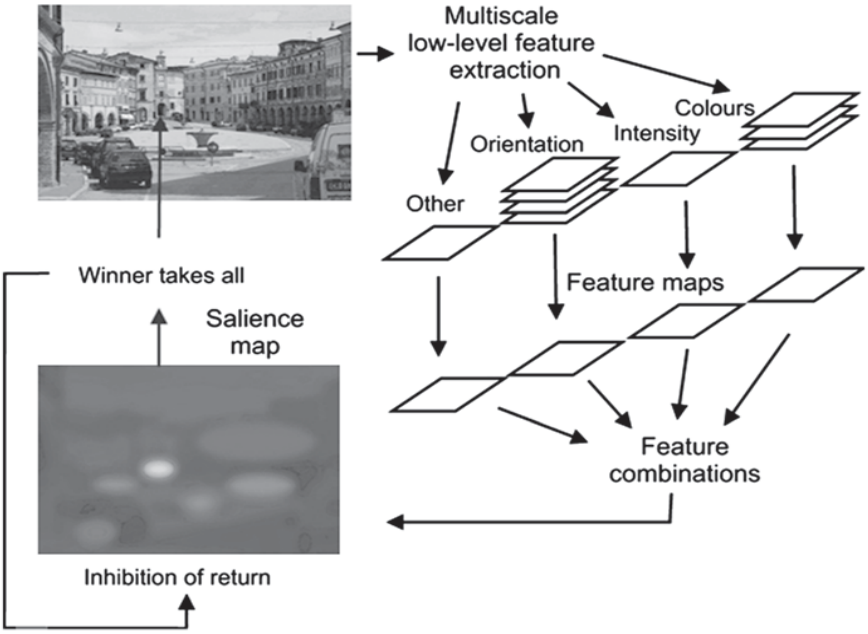


Figure 1.4. Based on Itti & Kock (2001) illustration of the organisation of a saliency map, reproduced from Land and Tatler (2009).

Torralba et al., (2006) proposed a contextual guidance model of attention in order to guide eye movements, combining bottom-up saliency, scene context, and top-down mechanisms (prior knowledge of where particular objects are likely to be found in a scene) which predicted the image regions likely to be fixated when performing natural search tasks in real-world scenes. Their model revealed the robustness of global contextual information in predicting observers' eye movements, and demonstrated the use of contextual guidance to constrain searches to areas based on gist representation and from our experience of where we are likely to find particular objects. Kanan et al., (2009) and Ehinger et al., (2009) expanded the contextual guidance model to try to predict where people look in a pedestrian search task; their model combined low level saliency, target features and scene context and found that when all of these features were combined the model could fairly accurately predict human eye movements. In addition to using spatial expectations to refine the search space in a scene, prior knowledge of the appearance of objects of a particular class can be used (Kanan et al., 2009).

Other authors have argued against the inclusion of saliency in a model predicting where we look and have demonstrated that models with no consideration of low-level saliency factors do just as good, or in many cases better jobs of predicting gaze allocation. Zelinsky, Zhang, Yum Chen and Samaras (2005) examined the contribution of saliency by computing a bottom up saliency map. They ran five combinations of two maps and found that the best performing model was one that had no contribution from the raw saliency map at all.

Several studies have examined the idea that our visual system is driven by reward (Sprague et al., 2007; Ballard & Hayhoe, 2009; Ropthkopf & Ballard, 2009; Ropthkopf et al., 2007). This type of model assumes that visual computations required in the real world can be broken down into a set of subtasks which is each associated with some reward value, specifically in the form of reducing uncertainty about aspects of our surroundings - and we therefore allocate our gaze based on the expected reward of fixating targets that reduce our uncertainty about unattended tasks. So for example in the experimental paradigm used in this instance of walking down the street, whilst performing the subtasks of avoiding static objects and picking something up, we allocate our attention based on reducing uncertainty, so would look at the pavement to reduce uncertainty about our direction, then after a certain point when that reward has been met, we can re-allocate our gaze to another area in the environment about which we have uncertainty, for example an object in our path, which then rewards us by reducing our uncertainty about the particular task of avoiding the object in the way. Thus where we look is based both on the task we are completing and the associated rewards of reducing uncertainty for all of the sub-task components that have to be met to reach the main task goal. In terms of reducing uncertainty, Renninger, Verghese and Coughlan, (2007) and Raj, Geisler, Frazor, and Bovik (2005) also argued that the visual system selects locations which give the most information for a task or which reduce uncertainty. They found that a model based on these factors was a better fit of human behaviour than the Itti and Koch (2000) model.

Further to the argument that other factors explain fixation selections better than low-level scene fixtures Henderson, Brockmole and Mack (2007) suggest that saliency may play no part in where we look, rather our visual gaze is driven by our expectations about congruency and this is what determines regions of interest and target fixation. Einhauser, Spain and Perona (2008) examined whether objects could predict gaze better than features and found that objects predicted over 65% of gaze whereas features performed at less than 60%. Further support for the importance of objects in visual guidance was found by Nuthmann and Henderson (2010) who demonstrated that the preferred saccadic landing position is close to the centre of the object. Foulsham and Kingstone, (2013) expanded this finding to demonstrate that this preference for the saccade to land close to the centre occurs from the first fixation of the object and even for relatively large saccades.

These pieces of work are clear demonstrations of how high-level factors such as task informs our visual behaviour, whether it is in terms of guiding our fixations to objects that are informative or in order to reduce our uncertainty about informative areas for task completion. However, with the exception of some of the studies examining reward (Sprague et al, 2007; Ballard & Hayhoe, 2009; Ropthkopf & Ballard, 2009; Ropthkopf et al., 2007) most of these studies, and the models proposed, consider only static scenes. Typically the types of tasks we engage in on a daily basis involve some form of action on our part and so, although the defining areas of interest are still set by task they are often also much more complex than those involved in simply viewing scenes.

One of the complexities imposed on viewing behaviour by active natural tasks is the element of sequential actions. Many of the actions we make in real life are made up of many small elements that combine to make an overall task goal.

Land and Furneaux (1997) note that lab based tasks are usually interested in the position of our fixations (i.e. ‘where we look’) in terms of reactive saccades, whereas in normal life our proactive saccades are formulated on a complex set of inputs, including predicting the physical laws of certain objects and actions, locating and recognizing informative areas and in dynamic environments tracking this, incredibly task-specific instructions, pattern recognition and memory for position information. Tatler (2009) pointed out that one difficulty with many of the proposed models is that they focus only on static scenes and thus treat all fixations as equal, whereas in the real world, particularly in the context of sequential tasks, we need to take into account fixation durations and where the previous fixation was. The intricacies of vision for action mean that where we look in natural environments during the completion of active tasks may be very different to where we look in scenes.

Where we look in the context of natural behaviour has been studied by several researchers in several different natural tasks, such as domestic tasks (Hayhoe, 2000; Land et al., 1999), driving (Land, 1992; Land & Tatler, 2001; Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003) and sport (Hayhoe et al., 2012; Land & McLeod, 2000; McKinney, Chajka, & Hayhoe, 2008). Just as other more static viewing paradigms have found that we fixate the regions that provide us with the information we require, the same has been found for our eye guidance during active natural tasks. Land & Lee, (1994) demonstrated that the fixations made driving along a winding, one way, single lane road showed a clear relationship between direction of gaze and steering and found that a considerable amount of time was spent looking at the ‘tangent point’ (the highly visible point on the inside of a bend where the drivers line of sight is lateral to the road edge) on upcoming bends. Drivers were able to look to other points as necessary and remain on course by making short gaze saccades, however these were made much less frequently as the driver was approaching a new bend, during which time 80% of fixation time was spent looking at the tangent point.

Driving is a learned task that requires constant visual input and monitoring. The findings from Land and Lee (1999) again point to the importance of task and directing our gaze to the most informative area, and although the ultimate goal in driving (i.e. to get from A to B) can be planned, the task itself has to be much more reactive since driving occurs in such a dynamic environment. Many other tasks performed in daily life are typically much more sequential and predictable in nature and allow for more detailed planning that can almost ‘script’ our actions. Domestic tasks are a good example of this and eye movements during these types of tasks have been studied by both Land, Mennie and Rusted (1999) who measured eye movements during tea making, Hayhoe (2000) and Hayhoe, Hayhoe, Shrivastava, Mruczek, and Pelz (2003) who studied eye movements during sandwich making. Our visual guidance in tasks that are sequential in nature is interesting since there is typically a diverse range of objects involved in the task, a variety of manipulations to be performed on the objects and each main goal is made up of many inter-related sub goals that must be completed sequentially in order to achieve the main task goal.

Both Land et al. (1999) and Hayhoe (2000) found that most of the time we look at objects that are relevant for the task, even although there are many other objects around that could potentially catch the eye. Hayhoe demonstrated that before the task of making a sandwich started fixations to irrelevant objects happened equally to looks to task relevant objects, however after the task began, fixations to irrelevant objects fell from 48% to 16%. Thus just as was initially suggested by Yarbus (1967) and Buswell (1935) task directly effects what we look at and can essentially constrain attention to only task relevant objects.

Although so far we have stressed the importance of looking at information-rich areas of a scene or environment, interestingly, there is a set of fixations which are actually directed to visually unremarkable areas of the scene, typically these fixations are to areas where an object is about to be set down (Land & Hayhoe, 2001). It may be that these fixations are simply the eye leading the hand 'guiding' the putdown of the object however it also reveals something about us predicting an upcoming event and directing our eyes to an area where something has yet to happen. Being able to predict where information is going to appear in a scene or environment, demonstrates that we build, and are able to act on, representations of scenes, environments and actions in order to formulate expectations and predictions. Furthermore, since setting down points are typically visually unremarkable and usually contain none of the low-level salient features that some models state are predictive of fixations, it is strong evidence for the argument that the eyes are driven by top-down mechanisms.

Predicting an area to fixate where information is about to appear is also witnessed in sports. Land and McLeod (2000) measured the gaze of cricket batsmen and revealed that rather than watching the ball as we might intuitively believe, the batsmen were able to anticipate the bounce point with gaze arriving

0.1 s or more before the ball. Land and McLeod argued that this demonstrates that looking at the bounce point provides the batsman with the information he needs to calculate where and when the ball will arrive at his bat, thus allowing him to work out his shot. The fact that gaze arrives to the bounce point before the ball maximizes the chance of acquiring this crucial information, particularly important in such a time pressured situation. A similar anticipatory fixation is made in table tennis (Ripoll et al., 1987) and squash (Hayhoe, McKinney, Chajka, & Pelz, 2012). In squash the eyes move to anticipate where the ball will contact the wall ahead of the ball by an average of 153 ms then rather than relocate to the second wall bounce point fixate on the ball's future trajectory after it bounces. The visual system is therefore able to make complex calculations based on our experience with the physical properties of the world in order to maximize our information gathering opportunities through the use of predicting up-and-coming areas of informativeness. Making predictions and anticipatory eye movements is revealing in terms of *where* we look but it is also intricately tied to *when* we look. The temporal nature of our visual behaviour will be discussed in detail in the following section.

1.5 When do we move our eyes

A considerable amount of research has been devoted to the subject of where we look in a scene, but not so much on the issue of when we move our eyes. The variability in the distribution of latencies (the time between presentation of the stimulus, and the response) poses interesting questions about the underlying processes that are occurring during this period. From the literature that does exist on fixation durations, several theories have been proposed as to what it is that is happening during fixations which causes differences in durations. Firstly it is proposed that the duration of fixations represents the visual and cognitive processes occurring during the fixation. This school of thought suggests that we terminate a fixation and initiate another saccade based on the current visual analysis of a scene (Henderson & Ferreira, 1990; Morrison, 1984; Rayner, 2009). The second process that is argued as underlying the durations of fixations is the notion of fixations being set to either an internal timer designed to keep the eyes moving or a timer, set at the beginning of viewing, which programs the duration of fixations depending on the nature of the scene and the task in hand (Henderson & Smith, 2009; Henderson, Weeks, & Hollingworth, 1999; Hooge, Over, Vlaskamp, & Erkelens, 2007). Alternatively, it is plausible that fixation durations are determined by a combination of the first two points (Henderson & Ferreira, 1990; John M. Henderson & Smith, 2009; Morrison, 1984; Yang & McConkie, 2001).

Carpenter's LATER (Linear Approach to Threshold with Ergodic Rate) model (Carpenter & Williams, 1995) explains latency distributions in terms of an underlying, essentially Bayesian, decision mechanism. The model proposes that the eyes move after using peripheral information to consider two alternative hypotheses, for example about the presence or absence of an object, gathering enough information to support the correct hypothesis (reaching a threshold/critical level) and guide the appropriate saccade. When enough information has been gathered, a saccade is initiated and in effect then the duration of a fixation is determined by the point at which the threshold is reached.

Moving the eyes when a certain reward threshold has been met is a theory furthered in a recent paper by Bray and Carpenter (2015). Which argues that the timing of saccades is considered to be driven by reward, that is, the reward of *amount of information* gathered. The authors demonstrate that saccadic latencies are shorter to a region that had previously provided reliable information than those that provided no information. Thus *when* we move our eyes may depend on what we expect to find there, and the speed at which we do this is influenced by the level of reliability we presume for the target.

The temporal nature of our eye movements is made up of the timing of saccades as discussed above, but also the length of time we spend fixating on an object, i.e. the fixation duration. One model, which proposes to explain the mechanisms behind fixation durations, is the CRISP model (Nuthmann, Smith, Engbert, & Henderson, 2010), that proposes a random walk saccadic timer that accumulates signal over time to reach a threshold. This can be modulated at any point by visual-cognitive effect, so as processing demands increase the speed of reaching threshold slows the saccadic time, which then leads to longer fixation durations. Once at threshold, a new saccade is programmed. The CRISP model asserts that the programming is a two-step process, which in the initial stage can be cancelled (labile stage), however the saccade programming during the following stage can no longer be cancelled (the non-labile stage). Although the CRISP model states that fixation duration does not equate to saccadic latency, nor does the time interval between two commands to initiate a saccade convert directly to fixation duration, the authors do argue that fixation durations at least partly reflect moment-to-moment cognitive processes during scene viewing. Accordingly the authors explain variations in fixation duration during different viewing task types as being due to a task-specific set of model parameters.

Several studies have examined the effect of task type on fixation durations, and revealed that individual fixations are typically shorter during visual search than memorization tasks (Henderson, Weeks, & Hollingworth, 1999; Võ & Henderson, 2009). Similarly object and scene semantics have been shown to influence fixation durations, with longer fixations on semantically informative objects (De Graef, Christiaens, & D'Ydewalle, 1990; Henderson et al., 1999; Underwood & Foulsham, 2006; Võ & Henderson, 2009). Thus *when* we move our eyes may also be influenced by the type of task we are completing, and informative properties of the scene.

When we move our eyes can be related to fixation durations but it is also worthwhile considering the sequential pattern of our fixations and how they relate to each other across time. Noton & Stark (1971) used the term scanpath to refer to a fixed pattern of eye movements that are performed repeatedly by the eye as a scene is viewed. Noton and Stark found that when viewing a simple line drawing, participants tended to repeat a sequence of eye movements during both a learning phase and then again for the first few eye movements when the pattern was re-presented. Noton and Stark did point out that, not only were the eye movements/scan paths different for different patterns when viewed by the same person, but different viewers displayed different gaze paths for the same pattern. Based on this they developed a "scanpath theory" which proposes that visual features are encoded and stored alongside a motor memory of the scanpath made during perception, Scanpath theory suggests that repeating a sequence of eye movements should facilitate memory. However Foulsham and Kingstone (2013) found no evidence that scanpath sequences made by a viewer are stored in long-term memory. Examining sequences of eye movements as scanpaths tells us something about the temporal order of fixations in space, but Foulsham and Kingstone (2013) point out that viewers tend to make fixations to areas in the scene which are informative and will aid future recognition of the scene and that any benefits as a result of this should not be attributed to any detailed representation of the actual sequence of fixations a viewer made.

In some cases, we may move our eyes when we have taken in and used the information we need from the current fixation and/or depending on some internal timer mechanism, however it is also interesting to consider instances where we look at an area in a scene which doesn't yet have any information present. In order to accurately predict up-coming visual information, we not only have to have an assumption of *where* it will happen we also crucially have to know something about *when* that information will be available in order to get our eyes there first. Experience allows us to formulate expectations and from these we can make predictions about when to direct our gaze to the next area likely to contain visually informative content in the near future. This ability to make predictions is not only seen in where we fixate but can also be demonstrated in our other visual behaviour, for example in smooth pursuit. We can formulate expectations about the future motion of a target and anticipate the motion with our eye movements. These anticipatory smooth eye movements were first noted by Dodge (Dodge, Travis, & Fox, 1930; Dodge, 1931) and subsequently reported by Westheimer (1954), who demonstrated that the eye would often turn around before the target during pursuit of a target, referred to as ‘anticipatory reversals’. Both Dodge and Westheimer claimed that these anticipatory reversals were due to learning. It could be argued that these eye movements are simply automatic and involuntary repetitious movements, some sort of low-level reflexive response to the stimuli, however Kowler and colleagues (Kowler, Martinc, & Pavel, 1984; Kowler, 1989) conducted an experiment which presented a target moving down either branch of an inverted Y-shaped tube with equal probability, sometimes the path was cued before the trial. The results showed that in instances where the path was not aurally cued, the past history of the target motion determined the subsequent velocity of the anticipatory pursuit. Whereas when the cues were present, anticipatory pursuits were based on the cued information. The authors argued that this is evidence that effects of the past were clearly overwritten by cognitive expectations about future events, thus the system is using high–level symbolic information to predictively guide the eye before any change in the actual stimuli. One recent study extinguished a moving target during different parts of the trial revealing that we are capable of using our expectations to counter-act VOR and make smooth pursuit like eye movements to ‘track’ an object that is no longer even there (Ackerley & Barnes, 2011). Essentially subjects successfully used motion information acquired from a previous presentation to track the target from the start of unseen motion generating smooth gaze. Thus *when* we move our eyes is also based on experience and the expectations we formulate based on that experience.

It has been demonstrated from studies such as the block-copying task (Ballard et al, 1992; Hayhoe, 1998) that eye movements are tightly coupled with the task in hand, both spatially and temporally. Spatially, in that we look mostly at the objects we need to complete the task, and temporally, in that these looks are not in random sequence, they are tied to the point in time when we most need the information. In fact, in the block copying tasks, the eyes were shown to precede the motor act by a fraction of a second. The patterns of fixations revealed that in a complex sequential task the eyes take one step of information at a time depending upon which element of the task they are guiding at that moment in time. The eyes had to support both the physical picking up and putting down of the block and the information supply of block characteristics. The typical pattern of fixations and actions was, to fixate the block in the model area, to remember its colour, to fixate a block of the same colour in the resources area, to pick up the fixated block, to fixate the same block in the model area, to remember its relative location, to fixate the corresponding location in the model area, to move the block, and drop the block. The only time that gaze and hand coincided was for periods of about half a second before picking up and setting the block down. Considering the relatively contained task area, it is interesting to note that information was extracted in instalments, with fixations made as and when needed. Ballard, Hayhoe, and Pelz (1995) demonstrated that if participants were asked to complete the block-copy task whilst fixating in a central spot the task took three times as long. The authors originally coined this strategy as ‘do it where I’m looking’ and in the subsequent paper Ballard et al. (1995) proposed a second maxim of ‘just in time’ to describe the fixation immediately preceding the action as providing the information for that action. We can conclude from these studies that the eye is guiding action and that we move our eyes either when we have extracted the piece of information that we needed, or we have guided the action (or a portion of it) successfully. However as Land and Tatler (2009) point out the pattern of fixations and actions here are quite specific to the task type and in normal everyday activity, it may be the case that more complex information extractions are needed to support action.

The temporal nature of gaze allocation in sport has been investigated in particular for ball games. Visio-motor control must be fast and precise, for example in cricket or baseball this means temporal accuracy of a few milliseconds and spatial accuracy of a few centimetres (Regan, 1992; Watts, 1991). The ability to make anticipatory eye movements has been found to be an advantage in a simple ball bouncing/catching task (Hayhoe, Mennie, Gorgos, Semrau, & Sullivan, 2004). Similar to batsmen in cricket, catchers initially fixate the hands of the thrower, then saccade to the anticipated bounce point, and then pursue the ball until it is close to the hands. Average departure time of gaze from the hands of the thrower was 61 ms after the ball left the hands. Catchers anticipated the ball's bounce point by around 53ms before the actual bounce. Anticipation of bounce points have also been demonstrated in table tennis (Land & Furneaux, 1997) and squash (Hayhoe et al., 2012).

In the previous section we discussed how anticipation in cricket involves directing the eyes to a location where there is upcoming desirable information, by necessity for these fixations to be considered anticipatory, they also have to be timed to occur before the physical event has happened. Again, anticipatory eye movements in cricket provide a good example to consider what influences the temporal nature of fixations. McLeod (1987) argues that batsmen cannot make changes to their stroke within the last 200ms before contact, meaning that the information needed to initiate the correct stroke must be obtained by the time of bounce. Regan (1992) points out that fast bowlers deliveries of speed are incredibly fast, frequently the ball reaches the batsman only 0.4 s after it leaves the bowler's hand, so timing here is crucial. Land and McLeod, (2000) tested batsmen facing differing paced balls of varying lengths, the batsmen's gaze revealed that expertise allowed faster latencies of initial saccades, meaning that good batsmen could anticipate the bounce point but poor or non-batsmen could not. Thus it seems that the temporal allocation of gaze is also affected by the level of expertise as well as task and experience.

The temporal nature of fixation allocation is an issue that has been somewhat neglected in the literature considering the wealth of work done on the way we view the world, however this may in part be due to the weighting of scene viewing as the most used paradigm in eye movement work. Studies which examine visual behaviour in the context of active natural tasks by necessity reveal the importance of timing with regards to eye movements, since acting on an object typically requires the visual information to have been extracted before the action can occur. Thus action itself imposes a temporal nature to visual behaviour that is just as important as spatial visual behaviour for our attempts to understand how we use our eyes to navigate our daily lives.

1.6 The importance of action

Returning to an idea put forward in the overview of this chapter, if we consider Wolpert et al's. (2001) argument that the point of the brain is to support movement and the purpose of vision is to supply information to the brain, then to study vision in the absence of action could fail to capture the way vision actually works in supporting behaviour. It is rare in daily life that we are not intending to interact with, or are currently interacting with our surroundings. In natural behaviour, the eyes provide the information we need to locomote, to play sport, to search for objects we need and guide our interactions with the object, and to read and write. Land and Tatler (2009) consider the contrast of what is essentially the modus operandi of vision in life compared with the laboratory based experiments where participants view images on screens, and question whether the lab based conclusions apply to natural behaviour. One of the issues with studying vision without action is that there is a tendency to focus on the spatial element of our visual behaviour, when in fact there is both a spatial and temporal element to eye movements in the context of active tasks which are both equally important. Traditionally, little work had been conducted examining both temporal and spatial characteristics of eye movements during natural tasks in extended environments (Canosa, 2009). However, advances in eye tracking equipment have meant that in recent years this has improved.

If the visual system is classified only as a mechanism which receives passive input and outputs the appropriate response then we would expect that regardless of one's intentions toward the visual input, our visual behaviour would remain fairly similar whether we were just looking at something or looking to interact with something. However, it has been demonstrated that there is a difference between vision for perception and vision for action, thus not all visual behaviour should be treated the same way and the study of vision for action is crucial in understanding human behaviour. Milner and Goodale (1995; Goodale & Milner, 1992) argued that the visual system originally evolved to enable animals to control their movements in a cluttered environment (i.e., action), not to provide more abstract knowledge of the world. Milner and Goodale proposed two neuroanatomically separate visual pathways in the primary cortex, the dorsal which supports online visual control of movements (the vision for action pathway) and the ventral which deals with information about objects, persons, events and environments (the vision for perception pathway). Part of the evidence for two separate visual pathways comes from individuals with optic ataxia, who can recognise properties of objects normally but have difficulty using visual information to control movements directed towards objects (Jakobson, Archibald, Carey, & Goodale, 1991; Jeannerod, 1986; Perenin & Vighetto, 1988). Similarly patients can display the reverse disassociation – visual agnosia, which presents with intact action but impaired perception (Goodale, Milner, Jakobson, & Carey, 1991). The double dissociation of visual agnosia and optic apraxia is strong evidence that we do indeed have separate visual pathways dedicated to support action (dorsal) and perception (ventral).

Further to the argument that vision for action is different to vision for perception, research involving visual illusions has provided evidence of different visual pathways. Van Doorn and Savelsbergh (2007) and Otto-De Haart, Carey, and Milne (1999) examined perception and action using the Müller-Lyer illusion. Participants were presented with shafts of differing lengths and found that the classic illusionary principles of the configuration held true, in that perceptual judgements of shaft lengths were affected by the orientation of the arrowheads surrounding the shaft. In contrast, the control of hand aperture when grasping the shaft was not affected by the visual context. Furthermore, significant differences in gaze patterns were revealed between the two tasks. More time was spent looking at areas that contained egocentric information (i.e., centre of the shaft) when grasping as compared to making a manual length estimate and also made more gaze shifts (i.e., mainly between the two areas surrounding the shaft endpoints and including the arrowheads) when making the manual length estimate, particularly during task execution as compared to task. The authors propose that the gaze shifts to the shaft endpoints were essentially enabling the extraction of allocentric information. According to van Doorn, van der Kamp, de Wit, and Savelsbergh, (2009), these results support the argument that the functional distinction between the dorsal and ventral systems is not limited to the processing of information, but also incorporates the detection of information.

The type of information that one is required to extract, depending on the task one is completing, is important for the visual system and can accordingly alter our visual behaviour. Andrews and Coppola (1999) found that active visual tasks that didn’t require physical manipulation but were nonetheless task driven rather than ’free-viewing’ (for example counting coins), elicited shorter fixations and larger saccade amplitudes than passive free-viewing visual tasks. Furthermore Canosa (2009) conducted an extensive series of experiments where the conditions varied in both the physical constraints of the task condition (from sat stationary at a desk to locomoting and performing an active task in the natural extended environment) and the task type and conditions (for example a block building copy task where the model was too far to visually refer to with ease). The results showed for example, that for tasks where more intricate manipulations to objects were required on the workspace (card sorting), more fixations to the workspace were made whereas for a block copy task with the model in plain view, equal proportions of fixations were made to both the materials and the workspace. This series of experiments demonstrated that not only do task demands change our behaviour but action itself imposes a set of task demands and therefore the characteristics of our eye movements during action are different from when we are looking for perception only.

In a clear demonstration of comparing how action changes our visual behaviour, Epelboim and colleagues (Epelboim et al., 1995, 1997) compared how subjects looked at a series of objects in two separate conditions. Participants had to either only look at a series of targets or had to tap a sequence of targets as rapidly as possible without making any errors. The results demonstrated that the way the oculomotor system samples the world differently depending on whether there are actions being carried out or we are simply looking for looking sake. Epelboim at al found that it took less time for participants to complete the task in which tapping was required, microsaccades were exceptionally rare when the head was free to move and the head was most likely to move before or at the same time as the eye, whereas when the participant was looking only, little head movement was observed. Supplementary evidence to this was provided by Bekkering & Neggers (2002) who used a visual search paradigm where subjects had to either look and point at a target or to look and grasp the target. They found that participants made fewer saccades to objects with the wrong orientation in the grasping condition than the pointing condition, whereas when the object was the incorrect colour, saccades were the same for both conditions. The authors argue that since saccade latencies were not different for the two conditions, no speed-accuracy trade off can explain the results, therefore the results suggest that specific action intentions (e.g. grasping) enhance visual processing of action-relevant features (e.g. orientation) supporting the argument that visual attention can be best understood as a selection-for-action mechanism.

As discussed in the previous section, studying gaze during sport is a useful tool to be able to examine the interplay of temporal and spatial demands on the visual system. Obviously sport by its very nature requires action, and in addition to revealing how visual behaviour generally guides action, Dicks, Button, and Davids (2010) compared gaze and movement behaviours of skilled football goalkeepers under two video simulation conditions (verbal and joystick movement responses) and three in situ conditions (verbal, simplified body movement, and interceptive response). They found that the goalkeepers spent more time fixating information from the penalty kick taker’s movements than ball location for all perceptual judgment conditions involving limited movement (i.e., verbal responses, joystick movement, and simplified body movement). Whereas, for the in situ interception condition when the goalkeepers were required to attempt to make penalty saves, an equal amount of time was spent fixating on the penalty taker’s relative motions and the ball location. The authors argue that their results suggest that gaze and movement behaviours function differently, depending on the experimental task constraints. These findings highlight the need for research on perceptual–motor behaviours to be conducted in representative experimental conditions.

Much of our daily life is composed of series of complex actions; typically we have a main goal, which is made up of several smaller sub-tasks which all have to be completed sequentially in order to achieve the main goal. Land et al. (1999) defined these sub-tasks as irreducible units of action sequence which involve the co-ordination of vision and action to carry out a manipulation of an object. Referred to as object related actions (ORAs) the eye movements that make up part of an ORA are typically characterized by the eyes fixating an object, or the point at which the object's activity is directed, *before* the manipulation starts (usually by about half a second). During a single ORA, multiple fixations will be made to the object being manipulated, whereas when shifting between ORAs, large saccadic amplitudes are often utilized to fixate on the next object in the sequence. These large saccades are a particular consequence of the task involving action (especially locomoting) and again highlight the importance of conducting eye movement research in settings that acknowledge the importance of action on our visual behaviour.

1.7 The importance of context

The study of human cognition has historically taken place in laboratory settings, focusing on the types of eye movements that can be considered visual sub-systems (Carpenter, 1988, 1991). So for example saccades or reflex adjustments and/or vergence movements will typically be studied in isolation of the other components that make up visual behaviour, which, whilst affording substantial experimental control, may also lose something of the holistic nature of visual behaviour in a more global sense. In some part this has been due to technological constraints but also largely due to the motivation for experimental control in order to discover causal relationships and therefore produce theories which are applicable universally (Kingstone, Smilek, & Eastwood, 2008).

In the early years of the study of eye movements (See Wade & Tatler, 2005 for a full historical perspective of modern eye movement research), recording devices were obtrusive, cumbersome and often painful to use, for example Ahrens, Delabarre and Huey in the late 1800's recorded eye movements using a plaster eye cup attached to a lever which would move a bristle at the end of a lever recording the eye movement onto a smoked drum of kymograph (Ahrens, 1891; Delabarre, 1898; Huey, 1908). Although this primitive eye tracker was unable to record high velocity saccades it is notable as the first actual method capable of recording of eye movements. However also notable was the serious drawback of causing mechanical damage of the eye (Eggert, 2007). Dodge and Cline, (1901) developed the first non-invasive photographic method and recorded the corneal reflection of a bright vertical line on a moving plate; this technique can be considered an early antecedent of the modern double Purkinje image (DPI) eye trackers. Subsequently corneal reflection is a feature of many modern eye trackers.

Initially trade-offs between accuracy and invasiveness were the biggest issues for earlier eye trackers, for example the search coil was very accurate but invasive, whereas the first infrared reflection device, developed by Torok, Guillemin, & Barnothy (1951) and subsequent IRD systems have proved difficult to ensure accuracy, the benefits of being non-invasive meant that they did become popular. Recently however, modern video-based eye movement recordings have become more popular perhaps this is somewhat attributable to the ease of use, affordability, improved accuracy, better temporal resolution and ability to cope with displacements between the head and eye.

Despite the technological advancements, even in the 1990’s much of the research was restricted to participants being stationary with headrests and chinstraps as a result of the type of recorders available. This research, despite any ecological limitations has been invaluable to our understanding of visual behaviour. On the other hand, evidence presented in previous sections of this chapter highlights the importance of vision being studied not just as an active system but also within the context of tasks that include physical actions. Obviously, technological restrictions meant that for a long time this was not possible, however the emergence of head mounted eye trackers around 50 years ago (Mackworth & Thomas, 1962; Shackel, 1960) meant that vision could be studied in the context of action for the first time. This revolutionary step allowed eye movements to be recorded in real world situations where the participant was engaged in an actual active natural task. Work set outside of the lab has demonstrated the tight coupling of temporal and spatial characteristics of gaze behaviour and active task completion (Hayhoe, 2000; Land et al., 1999; Land & Hayhoe, 2001; Hayhoe et al., 2003) and several researchers have in recent years begun to discuss how important the context vision research is studied in is for the broader conclusions we can draw (Kowler, 1990; Land & Tatler 2009; Steinman, 2003). The eyes are active information gatherers and thus it seems disharmonious to study visual behaviour only in the context of more passive lab based tasks.

Kingstone, Smilek and Eastwood (2008) put forward a case in their cognitive ethology paper for studying cognition in the context of how people behave in their real-world environments before moving to the lab. The authors argued that historical emphasis on experimental control by minimizing the complexity of the environment was flawed and that over the years it became increasingly clear that many causal relationships were specific to the laboratory conditions demonstrating that cognitive processes critically depend on the specific situational context in which a subject is embedded. Assumptions of invariance and control have meant that according to Kingstone et al, many researchers either deny the problem, or ignore the problem and continue with invariance assumptions, with only a minority of investigators acknowledging the problems and modifying their approach. As an alternative, Kingstone et al propose an approach referred to as cognitive ethology, which states that the flow of research should be driven from real-world observations to the laboratory in order to test some hypotheses formulated from the real-world observations.

Empirically, the importance of context and the disparity of findings from traditional lab based experiments compared with real-world natural tasks have been demonstrated by several researchers. Foulsham, Walker, and Kingstone, (2011) compared eye movements from the perspective of a person walking through the university campus to those recorded of a person actually performing the same walk. The results revealed some consistencies in gaze allocation during the two conditions, such as both sets of participants spent much of the time looking at the centre of the visual field, the types of objects that were inspected were also quite similar in both real walking and in video watching, other pedestrians in the scene were also fixated often between both conditions. However, crucially differences between the viewing behaviours of the two conditions were also found. Participants in the walking condition selected objects with head movements rather than making the large saccades that the watching participants made, the authors take this to imply that the lab findings may not reflect the dynamics of gaze selection in the real world. Although the eye movements from both groups displayed a central fixation bias, the walkers tended to fixate slightly below the horizon whereas watchers looked slightly above, the authors speculate that this may have been due to the walkers being more concerned with items below the horizon which would have been more likely to interrupt walking or in fact it could be due to the fact that for watchers there was unpredictability in the changes of head direction. They point out that not being in control of the head and therefore the field of view may incur a loss of predictability about the way the head will move and therefore influence our visual behaviour, which is important for the implications we can draw for laboratory based watching paradigms of this sort.

There are many examples from life where the conditions of even the same task vary dramatically depending on the contextual characteristics; one of these which has been studied empirically in relation to our eye movements is that of driving. Sivak, Conn, and Olson, (1986) recorded eye movements during real world driving in heavy, low speed, stop-and-go traffic conditions and found that fixations were concentrated primarily in the area of the rear view window, this is on contrast to studies involving higher speed, free flowing traffic where there is much variance in the placement of fixations (Sivak, Post, Olson, & Donohue, 1981). In heavy slow but steadily moving traffic, Land & Tatler (2009) describe gaze as alternating between the car in front (35% of the time), the right side oncoming vehicles (42%), parked cars on the left (10%) whereas when stationary at traffic lights, 75% of the drivers fixations were to things on left and right side pavements (shops, and pedestrians) and tended to be around twice as long. Further to this Underwood, Chapman, Brocklehurst, Underwood, and

Crundall, (2003) recorded fixations for novice and experienced drivers along three types of road (rural, suburban & dual-carriageway) and found that different parts of the scene focussed the driver’s attention to different extents, for example on dual carriage ways mirror inspections increased, as did the behaviour of scanning in the horizontal plane, leading the authors to conclude that visual behaviour is characterized as being sensitive to the prevailing road conditions on different types of road. Clearly then we can begin to see that context even within the same task of driving can have large implications on where we look.

Kingstone and colleagues (2008) argue that only after conducting observational studies in the real world should controlled laboratory experiments investigating the observed principles be conducted. An example of where this has occurred in the literature is from Land and Horwood (1995) who, following on from a real world driving study (Land & Lee, 1994) used a driving simulator to examine the influence of cornering information on normal and abnormal driving performance by systematically removing corner information normally available in the real world and found that the distance view was used to estimate curvature whilst near regions were used to estimate position. Interestingly at higher speeds better performance was achieved when the two road segments were visible, whilst for slower speeds only the near segment is necessary. It is difficult to imagine how the idea for this lab study could have been conceived without the initial real world observations from the original real world study; this is an effective example of the arguments made by Kingstone and colleague’s cognitive ethology approach.

We do not only use our eyes to seek out information; we also use other's eyes as sources of information to aid communication and in some cases to follow another's gaze cues. The importance of the context within which this visual behaviour is being measured has been shown to matter greatly in social situations. Lab based studies have demonstrated that when viewing images of people, viewers preferentially fixate on a person's eyes, and also follow gaze cues. (Birmingham, Bishchof & Kingstone, 2009; Friesen & Kingstone, 1998; Riccardilli, Bricolo, Aglioti & Chelozz, 2002). Whereas in real life we tend to avoid looking at and following other's gaze compared with looking at the same person on a screen (Laidlaw, Foulsham, Kuhn & Kingstone, 2011). These contradictory findings between how we behave in the laboratory compared with behaviour in naturalistic settings highlight the importance of studying visual behaviour in the context of active natural environments.

1.8 The importance of learning

As noted previously, infants as young as 4 months old appear to have some knowledge about the physical properties of the world and set expectations about some of the characteristics of objects including, support (Hespos & Baillargeon, 2008), occlusion (Aguiar & Baillargeon, 2002), position, (Wang et al., 2005), size (Wilcox, 1999). However as Land and Tatler (2009) point out, being able to execute many of the physical tasks that adults take for granted as being simple, actually take a long time for children to master. For example, typically children begin to learn ball catching skills at around 2 years of age, however it is not until around 5 years of age before they can consistently catch a bounced ball (Guitteridge, 1939; McClenaghan & Gallahue, 1978; Roberton &

Halverson, 1984). Sequential tasks like tea-making require several operational systems to interact (schema, gaze, motor and vision) and complete each individual component of the task. According to Land and Tatler (2009) children do not have all of the necessary skills to make a cup of tea before the age of 10, they point out that although many of the elements of the task of tea making will have been learnt early on, such as grasping and picking up an object and others such as filling a kettle will have perhaps been learned in another simpler capacity, the whole sequence of tea making would probably be more likely to be not perfected until around 10 years of age. By adulthood, experience and competency at basic sequential tasks like making tea mean that most of these tasks are performed off-line (Land et al., 1999), requiring little or no supervisory attention (Norman & Shallice, 1986). However since tasks like making tea are learned over many years often by doing similar task elements in different context, we do not currently have a good idea of how our visual behaviour guides and supports this type of learning.

When acquiring new motor skills, knowledge of the world and the characteristics of how objects behave in the world assist learning. The role of gaze during the acquisition of a novel and challenging visuomotor task was studied by Sailer, Flanagan, and Johansson (2005). A rigid hand held tool with two rotatable cylindrical handles controlled a cursor on screen with which participants were instructed to hit as many successively displayed targets as possible. In one condition, in order to move the cursor up and down the screen, opposite rotational torque had to be applied to the handles, whereas to move the cursor laterally the handles were required to be pushed together, this mapping was reversed in another condition. The authors hypothesized that after the task was learned, gaze would shift to the new target and remain there until the cursor arrived, but that during the learning phase either tracking the cursor would be done in peripheral vision or alternatively gaze fixations would be directed to the cursor during leaning. The results demonstrated that learning typically happened in a three-phase manner. The exploratory stage consisted of poor cursor control, with gaze often appearing to pursue the cursor, fixating recent cursor positions. During the next phase, referred to as the skill acquisition phase, performance improved quite dramatically and gaze more often than not fixated either the cursor position or the upcoming position. In the final phase there was still a gentle improvement in performance but this phase was considered a skill refinement phase, here both saccade and cursor were launched simultaneously toward a new target and gaze remained at the target until the cursor arrived. These results demonstrate how the role of gaze varies demanding upon the stage of acquisition, in this case initially closely monitoring the cursor, then moving to anticipating the cursor then to moving straight to the target and only monitoring the cursor for the final guiding stage of hitting the target.

In the example above participants were obviously able to acquire the new skill fairly quickly in one sitting, however in daily life it is common to go through these learning phases over a much longer time frame. Typically when learning, for example, a new hobby such as a sport, or painting, or a new musical instrument, even the very basics will be learned over a period of weeks, months, even years depending on the amount of time spent on the task. Although no longitudinal study of eye movements during expertise acquisition in natural tasks have been conducted, studies comparing novices to experts do reveal some distinct differences. Krupinski et al. (2006) recorded the eye movements of light microscopist with varying degrees of expertise; student, resident and pathologist, and found that when reading a virtual microscope slide the eyes are very quickly attracted to regions of interest (ROIs) within the slide and that these ROIs are likely to contain diagnostic information. In a matter of seconds, critical decisions are made on the selection of ROIs for further examination at higher magnification. Fully trained pathologists spent significantly less time scanning virtual slides compared to pathology residents or medical students, but had relatively longer saccadic eye movements. On the other hand, the pathologists spent significantly more time than trainees dwelling on the 3 locations they subsequently chose for zooming. Unlike either the medical students or the residents, the pathologists frequently choose areas for viewing at higher magnification outside of areas of foveal vision. The main finding of this study, that a fully trained pathologist can complete the same task of selecting ROIs using far fewer saccadic eye movements and generally in a fraction of the time demonstrates that not only does performance increase over time with more experience; expertise changes and refines visual behaviour.

As explored earlier in this chapter, for example in sport such as cricket and squash (Hayhoe et al., 2012; Land & McLeod, 2000; McKinney, Chajka, & Hayhoe, 2008) experts appear to use their experience when selecting fixation locations and their visual behaviour tends to differ somewhat from novices. Visual behaviour during driving has also revealed differences between novice and experienced drivers, particularly in terms of the types of sequences of fixations, with novices preferring to fixate on the upcoming road regardless of the road type and making more stereotypical sequences than those made by experts (Crundall & Underwood, 2008). Land and Hughes (in Land, 1999) examined differences in gaze patterns between novice drivers and their instructor and found a number of differences in where the novices looked compared to the experienced driver, with novices tending to confine gaze to the upcoming road and restricting looks deviating from the road ahead. Gaze behaviour when turning corners was the most strikingly different between novices compared to experienced, in that the instructor typically directed gaze by as much as 50° into the bend, whereas the learners all kept their gaze strictly in line with the cars heading. Crucially by the fourth lesson two of the three learners had learned to anticipate and were displaying the same gaze behaviour during turning corners as the experienced diver.

Further to the idea of the effect of expertise on gaze behaviour during driving Land and Tatler (2001) examined the eye movements of a professional racing car driver and found that the tangent point was still important and fixated on for much of the time but that in this case, the tangent point was based on the racing line rather than the bend itself. The authors also found that when overtaking, looks alternated between the car in front and the gap to the left or right and if overtaking, on a bend to the tangent point, again providing evidence that we look to the most informative areas for the task in hand. Land and Tatler propose that although the driver will typically have learned the course very well, it is still necessary to fixate on the racing line tangent point in order to check the exactness of his line.

Expertise is often quantified by some external measure/standard such as qualification, performance standing or professional position (e.g. as in the

Krupiski et al, 2006 paper mentioned previously), however, several studies have demonstrated that in simple tasks, we can all become ‘experts’ fairly quickly. For example Sailer et al. (2005) showed that participants could acquire a new and challenging visuomotor task in around one 20 minute sitting, whilst Hayhoe et al. (2004) demonstrated that participants not only anticipated bounce points in ball catching task, when this timing of anticipation was manipulated with a bouncier ball (with a different bounce point) the results showed that the visual system was able to update after just three practices and returning to anticipatory eye movements.

The point at which something has become so learned it can be considered automated is unclear, Foerster, Carbone, Koesling, and Schneider, (2011) went some way to investigate the process of automization using a bi-manual, high speed sensorimotor task where participants were instructed to complete a speed-stacking cup task for 14 consecutive days (for 45 minutes per practice session). The authors found that although participants produced similar scan-path fixation patterns across training days, the eye-hand span did become shorter. Which was argued to be evidence that automization of a high-speed sensorimotor involves long-term memory (LTM) regulating attentional focus. In further support for the notion of a LTM based control of attention, the same authors (Foerster, Carbone, Koesling, & Schneider, 2012) tested the same participants who had been trained in cup stacking in the first experiment this time participants performed cup stacking both in normal lit conditions and in complete darkness. Even in the dark condition, there were positive eye-hand latencies, that is to say that the eye arrived first at the object, despite the lack of visual information which the authors argued implied either, use of another type of sensory information or memory information. One of the principal roles of memory might therefore be to reduce the need to search for information: actions can be planned based on our memory of the task.

Memory and learning have also been previously demonstrated to combine in visual search tasks. Typically in real life we perform many tasks in environments we already have experience with, having already learned things like the layout, and the location of objects, which intuitively one would think would aid visual search. We know from several studies that objects that are relevant to the task are remembered better than objects that are not task relevant (e.g., Castelhano & Henderson, 2005; Williams, Henderson, & Zacks, 2005) but also that even distractor objects, that were not required to be remembered for the task, are remembered above chance (Hollingworth, 2006) and participants perform recall better than chance even when they are not expecting to be asked to recall objects (Tatler & Tatler, 2013a). However, several scene viewing search studies have suggested that instead of relying on memory for object locations, participants search afresh each time they are require to locate an object (Oliva, Wolfe, & Arsenio, 2004; Võ & Wolfe, 2012; Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011). Therefore it may be a case of what the visual system *can* do and what it actually *does* do, alternatively it may be something to do with the way in which the information is being encoded in the first place and in fact, one recent study suggests that attention ought to be paid to task specificity in visual search paradigms. Võ & Wolfe, (2012) found search benefit for having fixated an object in a previous exposure to the scene, but *only* if the previous exposure was for the purpose of search. No benefit was found for subsequent search for that object if the object was previously fixated during a scene memorization task. Thus, the purpose of fixation on an object appears to modify the encoding and retention of information. Hollingworth, (2012) replicated the study conducted by Võ & Wolfe (2012) study, but instead found that fixating for the purpose of judging semantic relationships between objects and scenes did improve subsequent search performance and argued that task specificity in memory is due to task-related differences in fixation allocation, but once fixation is allocated to a target, memory representations are encoded regardless of task goals.

From the evidence in this section we can conclude that there are several features of our visual behaviour that can be impacted by memory or learning. In terms of where we look and when becoming more experienced with a task appears to significantly change the way we behave visually both through the course of learning a task Sailer et al. (2005) and in terms of our performance in tasks such as sport, driving and job performance whereby the task may be learned but level of expertise is correlated with experience. (Crundall & Underwood, 2008; Hayhoe et al., 2012; Krupinski et al., 2006; Land & McLeod, 2000; Land & Tatler, 2001; McKinney et al., 2008; Land & Hughes, in Land, 1999). In tasks where there is an additional time pressure to the task, even eye-hand latencies have been demonstrated to be affected by learning (Foerster et al., 2011a). Ultimately we can conclude from these findings that learning can change the way we visually behave and that there is a distinct difference in how we view something when it is novel compared to familiar. Studies exploring eye movements during the process of learning have revealed that visual behaviour changes across the course of learning a task, however typically the types of tasks that have been used in these experiments either have been lab and screen based and did not include a sequential element (Sailer et al., 2005) or are performed in the extended environment but have pre-defined sequence which must be performed in the same way each time (Foerster et al., 2011a). Accordingly, it would be interesting to investigate what happens during the process of acquiring familiarity when the task is performed across time in natural extended environments and includes sequential sub tasks that have flexibility in the order in which they are completed.

1.9 Aims and scope of thesis

Several factors, such as the type of task we are undertaking, the environment we are undertaking it in and the objects we are using, may influence the way we visually behave during completion of a task. The level of prior knowledge or familiarity we have with these factors may have a significant impact on the way we search for objects, the time it takes us to complete a task, the order we complete a task in and even the microstructure of the ORA. To date, most of the work examining the effect familiarity has on visual behaviour has been to compare individuals who are already familiar (experts), with individuals with less, or no, familiarity (Krupinski et al., 2006; Land & Furneaux, 1997; Land & McLeod, 2000; Murphy & Wright, 1984; Myles-Worsley, Johnston, & Simons, 1988; Shinoda, Hayhoe, & Shrivastava, 2001; Stainer, Anderson, & Denniss, 2015; Underwood et al., 2003; Ward & Williams, 2003; Williams, Ford, Eccles, & Ward, 2011). Of the work that has looked at the process of learning, most of it has either been conducted on screens (Brockmole, Castelhano, & Henderson, 2006; Brockmole & Henderson, 2006; Oliva et al., 2004; Võ & Wolfe, 2012) or although conducted in natural environments the focus has been learning a task rather than the environment (Foerster, Carbone, Koesling, & Schneider, 2012b; Rebecca M Foerster, Carbone, Koesling, & Schneider, 2011b).

Whilst there are many situations in daily life were we might have to learn a new task, it is also the case that we often have to work in a new environment. Although several screen-based studies, as discussed above, have examined the effect of familiarity with a scene, the way familiarity influences how we interact with an environment and objects during the completion of an active task has not been studied. We live in a dynamic and complex world and as such our visual behaviour is required to flexibly cope with changes and new situations. The aim of this thesis is to examine what factors influence visual behaviour during the completion of familiar task and which of the fundamental principles of vision for action change, as we have to deal with objects and environments which we have varying levels of familiarity with. More specifically, we are interested in how sensitive our spatiotemporal allocation of gaze is depending on the context of the environment, the properties of objects and our level of prior knowledge.

This thesis will contribute to our understanding of vision for action and will further our knowledge of the factors that govern the spatiotemporal coordination of vision and action within complex real world activities.

Chapter 2 General Methods

This chapter outlines the methodology of all six experimental chapters of this thesis. In all of the following chapters the eye tracking methodology used was identical, wherever slight differences occurred, for example in calibration techniques, the specific details are discussed in each corresponding experimental chapter. Since all of the experiments for this thesis were conducted in natural environments using the portable eye tracker the resulting data to be analysed is initially captured in movie format with the fixation point superimposed onto the movie, therefore in order to quantify the visual behaviour during the task, videos are coded manually. Although there have been many studies completed using similar eye trackers which would have all required manual coding in the past, there is no agreed upon way to hand code data of this type. Deciding what to code in part depends upon the questions being asked and there are some more subjective elements of behaviour that require *a priori* rationale in order for coding consistency. The methods used and the reasoning behind coding decisions made will be covered in depth in this chapter.

2.1 Participants

All participants tested in the course of the following six experiments were recruited from a University student/staff population and therefore were between the ages of 18-35 and had a similar educational, socio-economic status. All participants reported normal or corrected-to-normal vision, soft contact lenses were permitted but data from individuals wearing glasses was discarded due to calibration difficulties. Informed consent was obtained from all participants before commencing and participants were made aware that they could withdraw their participation at any time with no penalty. Participants were either rewarded with course credits or were true volunteers. Participants were supplied with an information sheet outlining the purpose of the study but were all naïve to the purposes of the studies. All participants were provided with contact details for the lead researcher should they have any further questions.

In some cases participants participated in more than one study, all participants who took part in the experiment presented in Chapter 6 also deliberately took part in Chapter 7. Chapter 6 examines the changes in visual behaviour during the acquisition across ten days, whilst Chapter 7 probes what exactly was encoded during the acquisition of familiarity by presenting a two subsequent task switches, therefore all participants had had to take part in the familiarity acquisition phase. In a few other cases a couple of people also took part in two separate studies, the exact figures for instances where this occurred can be seen in Table 2.1 below.

Table 2.1 Percentage of participants taking part in one or more experiments

	Only this experiment	**2+ experiments**	
Chapter 1: Tea and Sandwiches	80%	20%	Chapter 4: Familiarity
Chapter 2: Glass Putdowns	100%		
Chapter 3: Kitchen Swap		100%	
Chapter 4: Familiarity		100%	Chapter 1: Tea and Sandwiches (100%) Chapter 5: Incidental encoding (20%)
Chapter 5: Incidental encoding		100%	Chapter 4: Familiarity Chapter 1: Tea and Sandwiches (100%)
Chapter 6: Follow up	50%	50%	Chapter 4: Familiarity

2.2 Stimuli

Since the primary focus of this thesis examines visual behaviour in the context of natural active tasks it was essential that all experimental settings and equipment were authentic. All experiments (except for Experiment 3) were conducted in the school of Psychology, Dundee. All experiments (except Experiment 2) were conducted in real, fully functioning kitchens using authentic kitchen equipment and paraphernalia. In Experiment 1, a staff kitchen in the school of Psychology at Dundee was used for all 40 sessions. In Experiments 3, 4, and 6, a lab kitchen was used to conduct all 133 sessions. Both kitchens were in light use by staff during the collection phases, however, it was always ensured that the kitchens contained the same equipment and that the set-up was identical for each trial.

The exceptions to these two kitchen set ups were Experiment 2 and 3. Experiment 2 was conducted in a lab using a bench set up, with real glasses, trays and jugs of water. The lab had ample space and clear bench tops for the purpose of the task. The justification for using a lab rather than a natural setting for this task being that the focus of this study was the objects rather than the environment. In contrast, Experiment 4 examined visual behaviour for a truly learned environment compared with an entirely novel one. In order to capture eye movements in an environment that participants were completely familiar with, data were collected in participants' own kitchens at home. As such, we had neither control over the task irrelevant objects present nor the layout of the kitchen, however the task remained the same and as such the objects used for each task were also kept consistent, even for studies conducted across multiple days

2.3 Eye Tracking Apparatus

All experiments used the same Positive Science LLC mobile eye tracker (New York, NY). The eye tracker consists of a lightweight glasses frame mounted with two small cameras (a scene camera and an infrared eye camera) tethered to small video recorders worn in a compact lumber pack. Also contained in the lumbar pack are two small batteries and a power pack. The scene and eye cameras record independently of each other to one of the dedicated video recorders contained in the lumbar pack. In order for optimal mobility the wires connecting the cameras from the glasses frame to the video cameras are worn to the participants back. The glasses frames are secured with elastic around the back of the head to ensure that as head movements occur, the cameras remain stable. The system records the eye using infrared light that is mounted on the glasses frame next to the eye camera, and tracks movements using pupil detection with an option to use corneal reflection data are recorded at 30Hz.

2.4 Calibration

Since the eye and scene cameras record independently of each other there is a possibility for temporal differences in recording, the Yarbus software allows for both videos to be synchronised together to correct for this by stretching or compressing the eye video time scale. To visually indicate the start and end points of both videos, a camera was flashed before and after calibration.

Calibration is manual and the points and planes to be calibrated to are set by the experimenter, for this thesis we calibrated each experiment to both a horizontal and vertical plane (worktop and wall), each experiment used dots set on paper as the calibration points, presented in a quincunx pattern. The experimenter is responsible for indicating which point the participant is to fixate at that moment in time, the dots were pointed to in a random order for approximately three seconds each and participants were instructed to look at each dot, without anticipating which dot to look at next.

Calibrating at both the beginning and end of task means that if during recording there are any tracking drifts the end calibration would be visibly inaccurate and either the data can be backwards calibrated if the point where the drift occurred or can be discarded if the track is no longer reliable.

2.5 Eye Tracking Software

When recording using the Positive Science eye tracker in the untethered mode, the output from each participant consists of two QuickTime movies, one of the right eye and one of the scene during the task. The Yarbus software supplied by Positive Science is compatible with the QuickTime movies and the synchronisation features allow both videos to be temporally matched up, correcting for any temporal differences in recording.

The eye position can be tracked using pupil thresholding and/or feature detection methods, with an option to use the corneal reflection. For the purpose of this thesis, the most reliable method of tracking was found to be feature detection with no corneal reflection however in one case (participant with certain eye features) the most accurate track came from using the pupil threshold. After synchronisation one file with the eye and scene movies together is presented which can then be calibrated. After which the files are rendered, a single video of eye tracking is produced showing the frame number, the scene, the eye picture-in-picture superimposed over the scene and the fixation points displayed either as a crosshair or a dot. The eye tracking movie is then saved as a quick time movie and the eye tracking data is saved as an ASCII text file, which provides the eye-in-head co-ordinates. This thesis used only the quick time movie files to code behaviour.

2.6 Coding

Coding for all studies in this thesis was completed manually; the total volume of data produced by the five experiments totalled around 1.4 million frames. The final eye tracking videos, produced using the Yarbus software, were observed using QuickTime, which allows for frame-by-frame scrolling. The data produced using the Yarbus software is a gaze cursor video which does not provide any data ready for analysis, similarly the ASCII files provide only the eye-in-head co-ordinates so there is a need to manually code the content of the videos. To do this the data were manually coded into excel spreadsheets in fields selected by the experimenter. When conducting manual coding there is no set way to code data, typically fixations and/or saccade amplitudes are measured but depending on the study and the purpose of the experiment there are many more visual behaviours that can be coded and measured when using a mobile eye tracker. Much of what is coded depends upon the questions being examined. Manually coding eye movement data can impose some risks regarding both the subjectivity of coding data into set categories, and in terms of reliability. For the present set of studies, in order to minimise these risks, strict criteria regarding the rules for coding and the rationale behind the rules were established prior to any raw data being coded, these rules minimised the subjectivity of making decisions moment-to-moment. Furthermore, since none of the studies in the present thesis had directional hypotheses, it would have been conceivable for most of the data to change in the opposite way as it actually did (or indeed to have remained unchanged) thus the risks again of both subjectivity and experimenter bias were minimised. One technique that can be used for future studies of this nature is to have a team of coders working on the same raw data, and then compare results to establish the inter-coder reliability. In the present thesis this was not possible since I was solely responsible for all of the coding and analysis, however the strict coding rules meant that subjective decision making was framed by detailed guidelines at the outset of coding thus ensuring consistency across all studies and data files. In instances where it is not possible to create such rationale coding criteria, another suitable option that could be utilised would be to code blind in order to minimise effects like experimenter bias.

The present set of experiments all examined the role of vision for action, therefore it was important to code what the eye did, what the hand did and some detail about the object being used and the action being performed. Whilst several of the studies were coded with exactly the same fields, two experiments were coded slightly differently since the focus of them was different, the details of specific coding for each experiment is detailed below. Two measures that were consistently recorded for all experiments (except Chapter 4) require further description. First, the latency between the eye arriving on an object and the hand arriving at the initiation of an action, this is referred to as the Eye-Hand Latency (EHL) as set out by Land and colleagues in the original tea making study (Land et al., 1999). The eye-hand latency is a period of time where vision leads action and has previously been assumed to occur with a latency of around a second (Hayhoe, 2000; Hayhoe et al., 2003; Land et al., 1999; Land & Hayhoe, 2001). Second, the measure referred to as look ahead fixations. In sequential tasks it is common for individuals to look ahead to objects that are not yet being used but will be in upcoming actions in order to achieve a sub goal. These types of fixations were noted by Pelz and Canosa (2001), during a hand washing task. Pelz and Conosa classified fixations as look ahead fixations if the look to a future use object occurred between around 500 – 1000 ms before the interaction with that object. More recently, look ahead fixations were looked at in further detail during the completion of a model building task (Mennie, Hayhoe, & Sullivan, 2007) classification of look ahead fixations in the model building task assigned any look to an object subsequently to be used within a 10 second time frame. In the present study, we included all looks to task relevant objects subsequently used (but not related to action at that point) as look ahead fixations.

2.6.1 Experiment 1

In Experiment 1 we were interested in the effect of movement and task on eye hand latencies, therefore the factors coded were as follows:

- The object being fixated
- The type of object (task relevant or irrelevant)
- The frame the eyes arrived on the object
- The frame the hand arrives on the object (if it was an object that included an action – some were simply looked at)
- The action being performed
- The frame the eyes leave the object

- The frame the hand leaves the object
- Which hand is used
- The type of putdown (guided or unguided)

2.6.2 Experiment 2

The main objective of this study was to establish whether the objects, and their specific properties, used in a task influence our visual behaviour and interactions with them as we put them down. Specifically we were interested in whether the inherent properties (like for example fragility, risk of tipping/breaking) determine the way we visually guide the setting down of objects with different 'risk' levels.

The experiment consisted of participants picking up and setting down several different types of glasses, some of these were filled with water to different levels and then set down on trays. We were interested in the eye movements during setting down of the glasses and trays of the glasses and manipulated the types of glass, the fullness of the glass and the number of glasses on each tray to be set down. The factors coded were as follows:

Table 2.2. Coded factors for glass properties

Coding factor	Options
Material	Glass
	Plastic
Glass type	Wine glass
	Champagne flute
	Tumbler

	High-ball
Fill level	Full Half full Empty
Number of glasses on tray	4 8
Type of putdown (see separate table for detail on putdown types)	Guided 1-7 Unguided 1-2

Results from Experiment 1 made it clear that when setting an object down on a surface, there were several ways to visually guide the action. Typically this has been coded in previous literature as either *guided* or *unguided,* however particularly for visually guided objects, setting down can be guided in a number of ways. Classifying the types of putdowns that were made for each object, in each manipulation, was attempted in the following way:

Table 2.3 Classifications of types of visual guidance used during putdowns

Guided 1	Looks at setting down location until put down is complete
Guided 2	Looks at setting down location until just before the object is set down then switches to object
Guided 3	Looks at object for entire setting down process
Guided 4	Looks at object until just before the object is set down, then switches to setting down location

Guided 5	Looks at setting down location then to another location/object (not relevant to current object)
Guided 6	Looks at object then to another location/object before putdown is complete
Guided 7	Looks at object for last few frames of set down
Unguided 1	No look to object or setting down place for entirety of putdown
Unguided 2	Very brief (3 frames or less) look to location or object at point of putdown, before fixating on another irrelevant object/location.

As indicated in Table 2.3 above, there are several ways vision can be used to guide the putting down of an object. The table above elaborates on the typical strategies used and reflects the level of detail data were coded to for each experiments in this thesis overall. In experiment 2 the put downs were in the first instance considered as either visually guided or unguided. For further analysis of the level of visual guidance deployed during the put down of glasses the different types of guided putdowns were grouped into those that had characteristics of continuous visual guidance (Guided 1 – 4 in table 2.3) and non-continuous (Guided 5 – 7). The Glasses were put down twice for each manipulation, once on a worktop surface before filling with water then once on the tray. The first putdown of the glass meant that the glass was always empty, whereas the second putdown on tray meant that the glass was always in one of the three *fullness* categories in the table above (empty/half-full/full).

2.6.3 Experiment 3

The focus of this study was to investigate whether prior knowledge of an environment influenced the way we visually behave in a well-learned environment compared to a novel one. The same factors coded in Experiment 1 were also coded in this, with the addition of noting the point at which there was a change in state regarding the object used in each action, i.e. if the action was to pick up an object, the change in state would be when the object was no longer in contact with the surface it was being picked up from. Both the start point of change of state and the end point were coded. The following nine fields were coded in excel spreadsheets for this study:

1) Code the object being fixated
2) The object is either:

1 - Tea
2 - Coffee
3 - Sandwich
D - Distractor
R - Room feature

3) The frame the eyes arrive on the object
4) The frame the hand arrives on the object
5) The action being undertaken
6) The frame a change of state occurs
7) The frame the action is complete
8) The frame the eyes leave the object
9) The frame the hand leaves the object

2.7 Rationale and details of coding

As mentioned in the beginning of this chapter, there is no set way to hand code data of this sort, therefor rationale for our coding decisions are set out in detail below for any of the above measures which are not self-explanatory, namely 5, 6, and 7.

2.7.1 The action being undertaken (coding rule 5).

Land et al (1999) described ORAs as irreducible units of action that when combined together and performed in a sequential manner, meet the overall goal of the task. The present study attempted to at all times code at the irreducible ORA level, therefore all sub-goals (for example ‘Switching kettle on’, ‘Picking up teacup’) were coded. There were also many instances where an object was fixated but not involved in any action. This would be coded for fixation but no hand or action data would be recorded so, the frame the eyes arrived and left but the frame the hand arrived and left would be blank.

2.7.2 Change of state beginning (coding rule 6)

Further to classifying the frame the eye and hand arrived/left and the action completed, the present study investigated the time taken to complete each action and the effect this had on visual behaviour. Therefore, we classified the action element of manipulations on objects as having Changes of State (COS), with each COS having both a start and end point. The reason this cannot be taken from a calculation of the frame the hand arrived to the frame the hand left is that it is common for either the hand to arrive before manipulations commence or for the hand to linger after an action has been completed, therefore coding the start and end point of an object’s state and the change in state is a more sensitive and precise measure for this type of investigation.

The actions involved in most COS start points in tea making are straightforward, for example picking objects up, pouring, stirring and opening lids/drawers/doors. If an object is picked up e.g. the teapot, the change of state is coded as the first moment any part of the teapot leaves contact with the worktop or if the action is a drawer being opened, the COS occurs with the first appearance of an opening between the drawer and the drawer frame. However, there are instances where this rationale is not as clear cut, for example when the action involves putting objects down and closing doors/drawers/lids since these actions are initiated before sometimes long before a change of state occurs. For example, if the action is ‘putting teapot on worktop’, the COS is coded as the frame where the first part of the teapot comes into contact with the worktop, since this is the first point at which there was a change of state. Although the put down of the teapot will be initiated before this, until this exact point the state of the teapot would still have to be considered as ‘up’, whereas, the moment the first part of the teapot comes into contact with the worktop, the state has changed. If we were to ask for the teapot to be moved, at that point we could say ‘pick up the teapot’ for the first time, thus there has been a change of state even if the action is not quite complete. The rationale for not coding from the initiation of the ‘putting down’ is twofold, firstly we have a rough measure of that from the hand arrived frame. Secondly, there is too much subjectivity in the exact point an action like this is initiated. For example, if one retrieves the teapot from a shelf up above the work surface and in a continuous movement puts it down on the work top, the point at which we could classify the start of the action ‘Puts down teapot on worktop’ is difficult to precisely determine.

Similarly with opening drawers and doors, the initial change of state for a drawer/door opening is the first point at which an opening is visible. For drawer/door closures, again the point where the closure was initiated was not the start point for the coding, since again there are instances where an incorrect door is opened and in one continuous movement is immediately closed again, therefore determining the beginning of the actual point where the drawer is closing is impossible in this case. Also, up until the drawer reaches the edge of the drawer frame, we would have to consider the state of the drawer as open. Therefore we code change of state from the point the drawer comes into contact with the drawer frame as this is the point the state has changed from open to closed (again, even if it has not finished closing completely).

2.7.3 Action complete (coding rule 7)

There are many instances where the action is complete but the hand lingers on the object for a few frames. Therefore when coding, for example the action ‘puts the teapot down on the worktop’, the action was coded as being complete when the teapot was at rest completely (which may or may not be several frames before the hand leaves).

Slightly more complicated are cases with doors and drawers opening. There is a point where the drawer/door is no longer in contact with the frame (normally just a few frames after the change of state) however this is typically not the end of the action ‘opens drawer’. The point at which the door is at rest and no longer being opened is the point at which the COS is complete.

2.7.4 Experiment 4 and 5

All coded in exactly the same way as Experiment 3, using the same rules and rationale.

2.8 Missing fields and data loss

The nature of mobile eye tracking in natural environments incurs the risk of some data forfeiture due to equipment failure, set-up errors or participant induced losses. Care was taken to successfully avoid either of the first two loss types in this thesis, however some individual data points were lost due to the limitations of the eye tracker scene camera’s field of view. For example, it was a relatively common occurrence for participants to look to the next object before having removed their hand from the current object under manipulation, so in this instance the *hand left* frame would be missing since it would not be in view of the scene camera and therefore would not be visible to code. Similarly, there were some instances where the participant could complete part of or all of an action off camera – that is to say that their gaze would be allocated to another object entirely and the object being manipulated would not be in view of the scene camera, these instances were obviously not included since no values were available for the eyes or hands during these manipulations.

One missing field, which did occur fairly frequently, was related to the specific action of picking up or setting down certain objects. In some instances the action to change the state of an object was so short that the COS beginning to COS end was the same frame, for example when the milk lid was lifted from the worktop, typically it would be complete in 1 frame. Instances such as these were coded as having only a COS beginning frame therefore providing a missing value in the COS end frame, these instances were still included in the analyses.

In Chapter 3 it is noted that data from 10 participants was deemed unusable, it is subsequently explained that this period of data collection occurred during the training phase of equipment use and as such all of the losses of data were due to experimenter error. Due to the between subjects design of the study, for a participant's data to be usable, all four sessions had to have been successfully recorded meaning that if an error occurred on just one session the entire participant was removed. After the initial data losses, further training and practice with the equipment was implemented, along with the procurement of spare batteries to power the eye-trackers, these steps ensured that no further data was lost for the following experiments in the thesis.

Chapter 3 Factors that influence the temporal relationship between vision and action

3.1 Introduction

Real world activities can be reduced to a set of component actions that are linked together in order to produce complex behaviours (Miller, Galanter & Pribram, 1960; Forde & Humphreys, 2002). Understanding the irreducible functional units of behaviour is an important first step in understanding complex human behaviour (Schwartz et al., 1991). In many everyday tasks, the irreducible behavioural unit involves the coordination of vision and action in order to achieve a visuomotor routine required for the current sub-goal, a behavioural unit referred to as the object-related act (ORA; Land et al., 1999). The microstructure of the ORA is surprisingly consistent across a range of real world activities, with both spatial and temporal coordination of vision and action. In space, central vision is directed to the target of the manipulation (Ballard et al., 1992; Land & Tatler, 2009). In time, central vision is directed to the target of the manipulation about 0.5 – 1 second before the hand makes contact (Land et al., 1999; Land & Tatler, 2009). Thus if we can understand the allocation of gaze in space and time around our actions, we may gain valuable insights into the organising principles for human behaviour.

Typically our day-to-day tasks are sequential in nature, with several interrelated but distinct actions, often requiring continuous uptake of visual information. This view of task structure, as a hierarchy of goals and sub-goals was formalized by Schwartz and colleagues (Schwartz et al., 1991, 1995) in a standardised action coding system (ACS), intended to be used for assessing brain-damaged patients' performance of everyday tasks. Land et al. (1999) extended this hierarchical view of task performance by suggesting that the irreducible unit of behaviour was the object related act (ORA), which involved the coordination of vision and action to carry out a manipulation of an object (Figure 1). Understanding gaze allocation in space and time within an ORA may therefore offer important new insights into not only the allocation of gaze in natural settings but also the manner in which more complex behaviours may be built up. To date, the factors that govern spatiotemporal coordination of vision and action within an ORA have not been investigated systematically. The present study takes this novel approach to understanding the link between vision and action in real world activities.

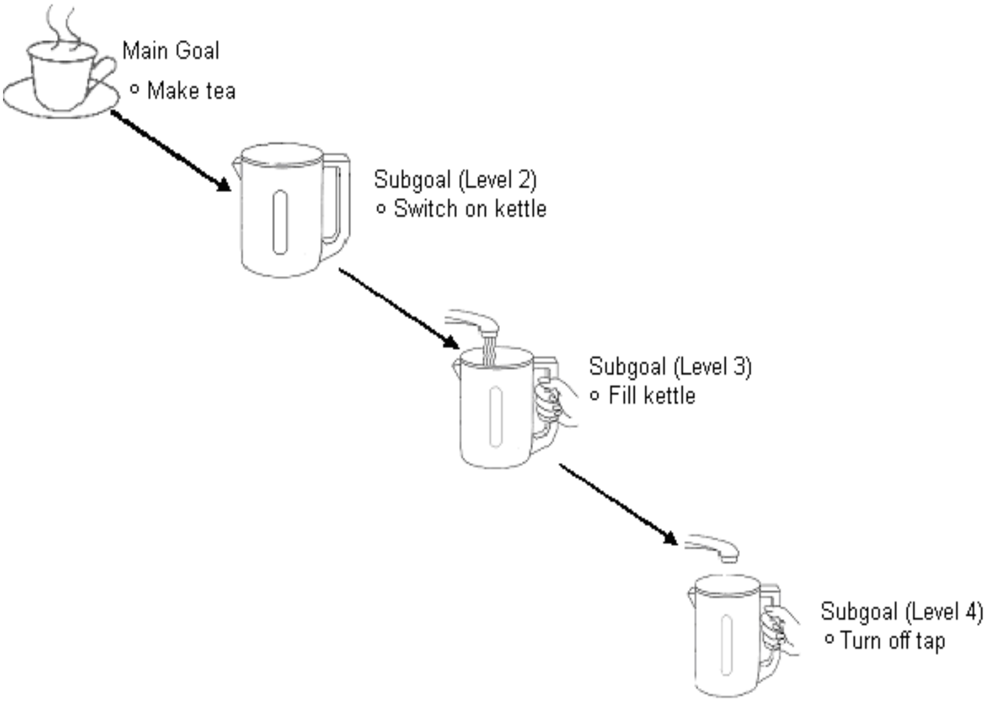


Figure 3.1 Levels of ORAs, Level 4 being the smallest irreducible level of action plus accompanying eye movement.

Object related acts typically require about 3 seconds for completion and involve an average of 5.4 fixations (Land et al., 1999). In terms of the spatial allocation of gaze within an ORA, most fixations are directed at the object that is the target of the current manipulation (Ballard et al., 1992; Land & Tatler, 2009). Despite there often being multiple fixations within an ORA, most remain on the target object, and it is not clear whether all of these relocations necessarily provide essential new information about the object (Tatler et al., 2011). Exceptions to this rule are fixations that disengage from the current act temporarily and fixate the target of a future act, before returning to the target of the current act, often referred to as “look-ahead” fixations (Mennie et al., 2007; Pelz & Canosa, 2001). Understanding spatial allocation of gaze within an ORA is therefore somewhat trivial: we look at the object that is the target of the ORA.

With regard to the temporal allocation of gaze within an ORA, vision is proactively allocated at the start and end of each ORA: when making tea, the eyes are directed to the target of the ORA about 0.5 – 1 seconds before the object is manipulated, (Land et al., 1999). This temporal relationship not only emphasises the proactive nature of gaze allocation in everyday activities but also highlights the importance of understanding gaze allocation in time with respect to our current actions.

The timing of gaze allocation with respect to action is surprisingly consistent across a wide range of real world activities. The eyes tend to lead motor output by about 0.5 – 1 second in activities as diverse as driving (Land & Lee, 1994; Land & Tatler, 2001), music sight-reading (Furneaux & Land, 1999), walking (Patla & Vickers, 2003), and reading aloud (Buswell, 1920). Furthermore, this temporal relationship between vision and action seems to develop as new visuomotor skills are acquired (Sailer et al., 2005). Thus understanding this consistent temporal link between vision and action may be a key step for furthering our understanding of the fundamental building blocks of behaviour.

The consistency in eye-hand latency found across the range of tasks mentioned above might suggest that this temporal relationship is constant across everyday activities. However, two studies of similar, domestic tasks raise the possibility that this relationship may not be as fixed as implied by the studies reviewed in previous paragraph. When making tea (Land et al., 1999) or making sandwiches (Hayhoe, 2000) the same principles of visuomotor coordination are evident: with few fixations of task-irrelevant objects, and close links between vision and action in space and time. However, while the eyes typically led the hand in both tasks, there was a considerable difference in the eye-hand latencies between the two studies. During the sandwich-making task, the average eye hand latency was 0.09 seconds compared with the much longer 0.56 seconds found in the tea making study. This raises the possibility that the temporal link between vision and action may vary under some circumstances. Land and Hayhoe (2001) speculated that the difference may arise from the setting in which the tasks were completed. Specifically, when making tea, participants were required to move around the environment to complete the task, whereas when making sandwiches participants were seated throughout, with all objects within reach. Land and Hayhoe (2001) suggested that the need to move around necessarily imposed a slower tempo for completing the tea making task, with more time between ORAs and therefore greater opportunity to fixate the next object earlier relative to contact by the hand.

An interesting case for the spatiotemporal allocation of gaze during an action is situations in which objects are placed on a surface. In some cases these placements are preceded by fixations to the empty space on the surface where the object will be placed (Land et al., 1999). These situations provide an interesting challenge for our understanding of gaze allocation in space and time because the associated fixations are directed to locations that are visually unremarkable when they are initially selected (see Tatler et al., 2011). The tendency to guide object placements also varied between tea making (95% of put-downs were guided) and sandwich making (87% of fixations were guided). Why guidance was less prevalent in the sandwich-making task is unclear, but Land and Hayhoe (2001) again speculated that being seated in the sandwich task might have been important. Land and Hayhoe did not consider whether the differences might in fact have been due to the different tasks performed in the two separate studies, presumably since both were domestic tasks.

In the present study I explored two possible factors that might influence the temporal relationship between vision and action and the tendency to visually guide object put-downs. First, whether the requirement to move around the environment to complete a task influences aspects of visuomotor coordination (a possibility suggested by Land & Hayhoe, 2001). Second, whether differences in the task goals themselves might be responsible for previous reports of differences in the spatiotemporal coordination of vision and action. The present study used a 2 x 2 within subjects design in which participants had to (over four testing sessions) make tea and make a peanut butter and jam sandwich, both in conditions where all objects were within reach and in conditions where objects were distributed across two regions of the environment such that the participant to walk around the kitchen in order to complete the task. This study offers the first systematic exploration of the factors that govern spatiotemporal coordination of vision and action within complex real world activities.

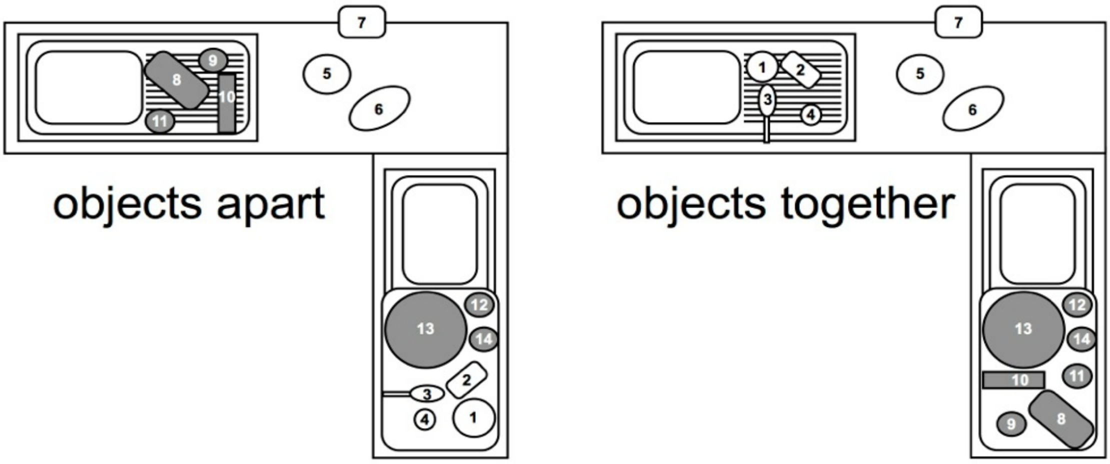
3.2 Method

3.2.1 Participants

20 (2 male) undergraduate psychology students from the University of Dundee participated in the study in return for course credits. All participants had normal or corrected-to-normal vision.

3.2.2 Materials

A fully functioning kitchen in the University of Dundee, School of Psychology building was utilised for this study. The room fixtures included worktops and kitchen sinks, shelves and electricity points. Task relevant objects, i.e. those for making tea and sandwiches were laid out in the kitchen along with several distractor objects (Figure 2). Distractor objects included several items typically found in the kitchen, for example a dish sponge, and also several items not typically associated with kitchens, for example a hand held fan. Perishable objects for both tasks were laid out in the kitchen and replaced as and when necessary.



object key	
1. teapot	8. bread
2. milk	9. jam
3. teaspoon	10. knife
4. sweetener	11. cola
5. mugs	12. glass
6. kettle	13. plate
7. tea	14. peanut butter

Figure 3.2 Layout of the kitchen for all conditions

3.2.3 Eye Movement Recording

Eye movements were recorded using the same equipment and procedure as outlined in Chapter 2. Frames from the gaze fitted videos of a participant completing both the sandwich-making task and the tea-making task are presented in Figure 3.3.

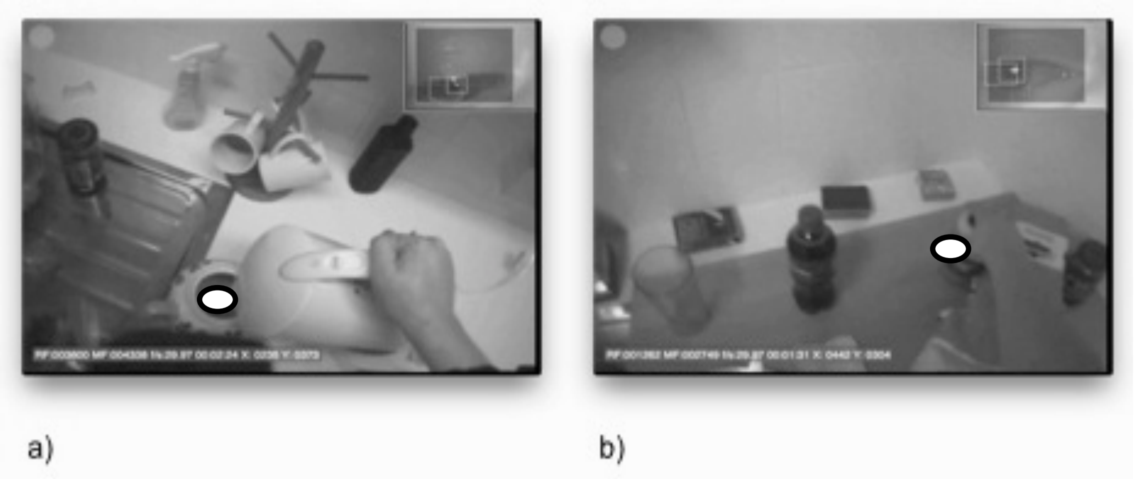


Figure 3.3a) A frame from the gaze-fitted video of a participant making tea in the objects-apart condition, b) a frame from the gaze-fitted video of a participant making a sandwich in the objects-together condition

3.2.4 Design

We used a within-subjects design to manipulate two independent variables (task and requirement to move). The two task conditions involved making a cup of tea and making a peanut butter and jam sandwich. We manipulated the requirement to move by placing all objects for a particular task in one area of the kitchen (requiring no movement during the task) or by spreading the objects over two areas of the kitchen that required the participant to move between these areas during the task (see Figure 2). When the objects were all in one area (the “objects together” condition), we emphasized to participants that they should stand in one location throughout the task and not walk around the kitchen. When the objects were spread across two areas (the “objects apart” condition), we explained to participants that they would need to walk around the kitchen to complete the task.

When making tea (in both the “objects together” and “objects apart” conditions) participants were asked to use the teapot and make tea with sweetener and milk. Instructions on the tasks were kept consistent across both the “objects together” and “objects apart” conditions, the only difference being whether participants were permitted to move around the room or not.

3.2.5 Analysis

Data were discarded for an entire participant if on any of their four recording sessions, the calibration procedure could not provide a reliable estimate of gaze position, or if any of the four recording sessions was not recorded properly (e.g., due to the batteries running out in the camcorders). After these strict exclusion criteria data from 10 participants were available for subsequent analyses (8 female). All of the data lost was due to experimenter error, data was collected during the training phase of mobile eye-tracking and as such errors in terms of setting up equipment correctly incurred the loss of data from 10 participants.

Gaze-fitted movies from each recording session were analysed manually on a frame-by-frame basis. For all eye movement related measures, we recorded gaze events rather than individual fixations: a gaze event was defined as the time from the first entry to an object to the first exit from that same object, irrespective of the number of fixation made within the object. For each manipulated object (thus for each ORA in the task) we coded the time that gaze was first directed to the object, and the time that the hand made contact with the object. Gazes made during the putdown of objects were also recorded; these were considered either as guided, if gaze was directed either to the object or the location the object was eventually set down on, or as unguided, where no visual guidance was used for the entire setting down process. Gaze events to task irrelevant objects were also coded, as were instances where an object later to be used was fixated in advance, without a related action.

Each DV was analysed using 2-way repeated measures ANOVAs with task (tea, sandwiches) and object layout (together, apart) as factors. Partial η^2 is reported as a measure of significant effect sizes. DVs were: the number of ORAs carried out to complete the task, the percentage of gaze events that were look-aheads, the percentage of gaze events directed to task-irrelevant objects, eye-hand latency at the start of an ORA, and the percentage of guided object put-downs. For eye-hand latency data, outliers were removed if they were more than 2.5 standard deviations from the mean for that experimental condition. We analysed mean eye-hand latencies for comparability to previous studies.

3.3 Results

3.3.1 Effects of task and layout on visual behaviour

The number of object related acts carried out by participants did not vary between the two tasks, $F\ (1, 9) = 1.41$, $p = .265$, or the two layouts of objects, $F\ (1, 9) = 2.22$, $p = .171$. There was no interaction between the task and layout, $F\ (1, 9) = 0.30$, $p = .598$ (tea, objects together = 25.3; tea, objects apart = 25.9; sandwich, objects together = 26.5; sandwich, objects apart = 29.2).

The number of look-ahead fixations varied between the two tasks, $F\ (1, 9) = 10.90$, $p = .009$, $\eta^2_{\text{partial}} = .548$, with 27.9% of gaze events being look-aheads when making tea and 18.1% of gaze events being look-aheads when making sandwiches. There was no main effect of object layout, *F* (1, 9) = 0.02, *p* = .902, nor did these factors interact, *F* (1, 9) = 1.08, *p* = .327.

The percentage of gaze events directed to objects irrelevant to the current task was greater when making tea (11.2%) than when making sandwiches (3.4%), *F* (1, 9) = 15.59, *p* = .003, $\eta^2_{partial}$ = .634. There was a main effect of object layout on the percentage of gaze events directed to task-irrelevant objects, *F* (1, 9) = 18.49, *p* = .002, $\eta^2_{partial}$ = .673, with more gazes to task-irrelevant objects when objects were distributed across two areas in the kitchen (11.3%) than when all objects for the task were together (3.3%). There was a tendency toward an interaction between task and object layout, but this failed to reach significance, *F* (1, 9) = 4.56, *p* = .062.

Table 3.1 Visual behaviours averaged across participants. Standard deviations in parentheses.

	Tea		Sandwich	
	Together	Apart	Together	Apart
Total Number of ORAs	25.3 (5.77)	25.9 (7.52)	26.5 (8.76)	29.2 (7.16)
Proportion of 'Look-Aheads'	.30 (.09)	.26 (.09)	.16 (.09)	.20 (.12)
Proportion of looks at task irrelevant objects	.05 (.07)	.17 (.12)	.02 (.02)	.05 (.04)

3.3.2

3.3.3 Eye-hand latencies at the start of an ORA

There was a main effect of task on the eye-hand latency at the start of an ORA, *F* (1, 9) = 43.23, *p* < .001, η^2_{partial} = .828, with longer eye-hand latencies when making tea (0.72 seconds) than when making sandwiches (0.57 seconds). There was no main effect of object layout on eye-hand latencies at the start of an ORA, *F* (1, 9) = 0.16, *p* = .702, nor was there an interaction, *F* (1, 9) = 0.09, *p* = .767.

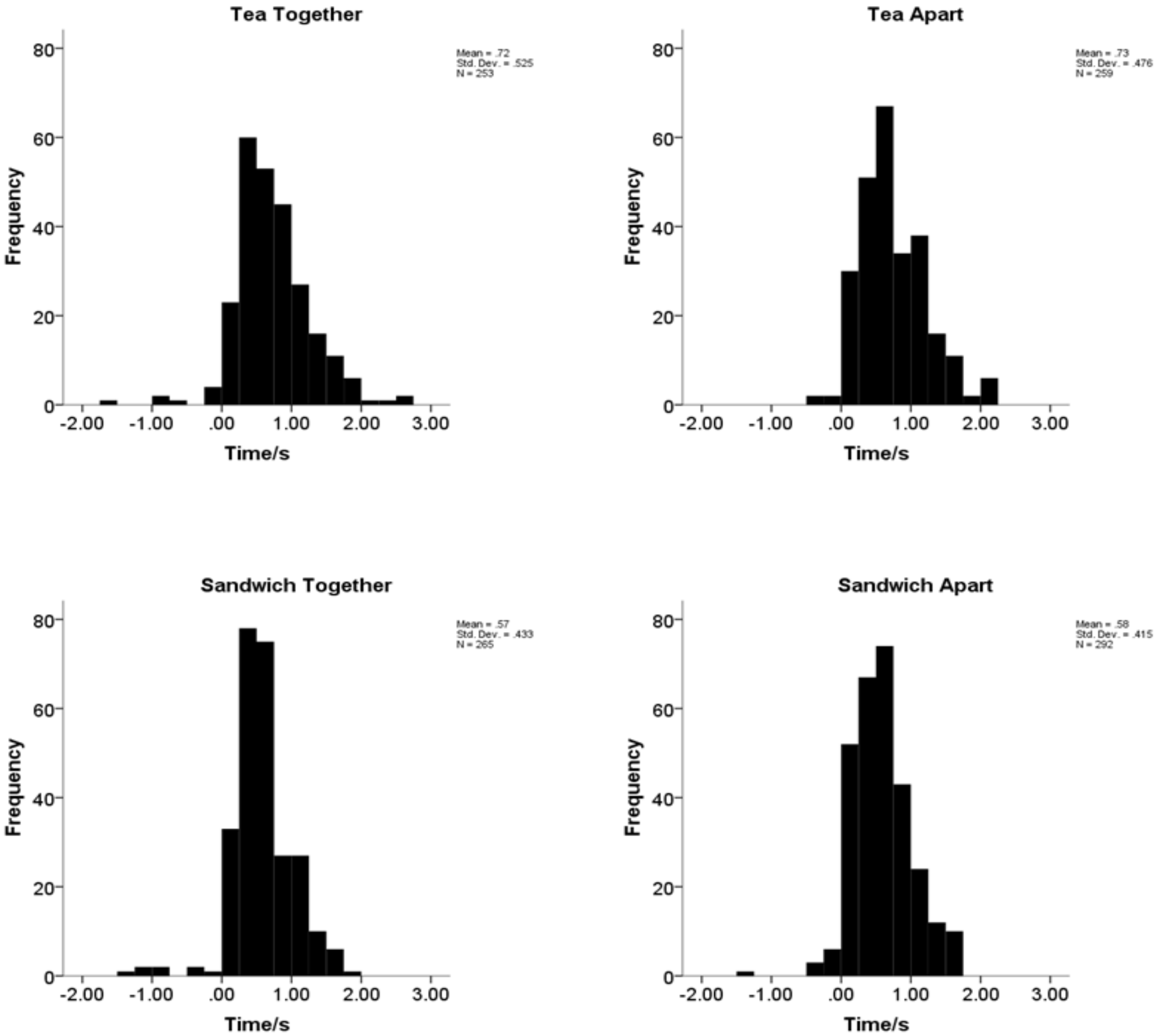


Figure 3.4 Eye hand latencies for beginning of ORAs across all four conditions

3.3.4 Visuomotor coordination during object put-downs

There was a main effect of task on the percentage of guided object put-downs, *F* (1, 9) = 7.17, *p* = .025, η^2_{partial} = .443, with more guided put-downs when making tea (64.8% of object put-downs) than when making sandwiches (48.3% of put-downs). There was a tendency toward more guided put-downs when objects are spread across two areas of the kitchen (59.8% of put-downs) than when all objects needed for the task are together (53.3% of put-downs) but this failed to reach significance, *F* (1, 9) = 3.63, *p* = .089. There was no interaction, *F* (1, 9) = 0.24, *p* = .633.

An ANOVA was run on a subset of the data with selected objects to examine the effect of object and movement condition on eye hand latency. As with the previous result, there was no significant effect of movement (*F* < 1), but there was an effect of object type (*F* (1,7) = 2.802, *p* = 0.005). There was no interaction between the two effects (*F* < 1).

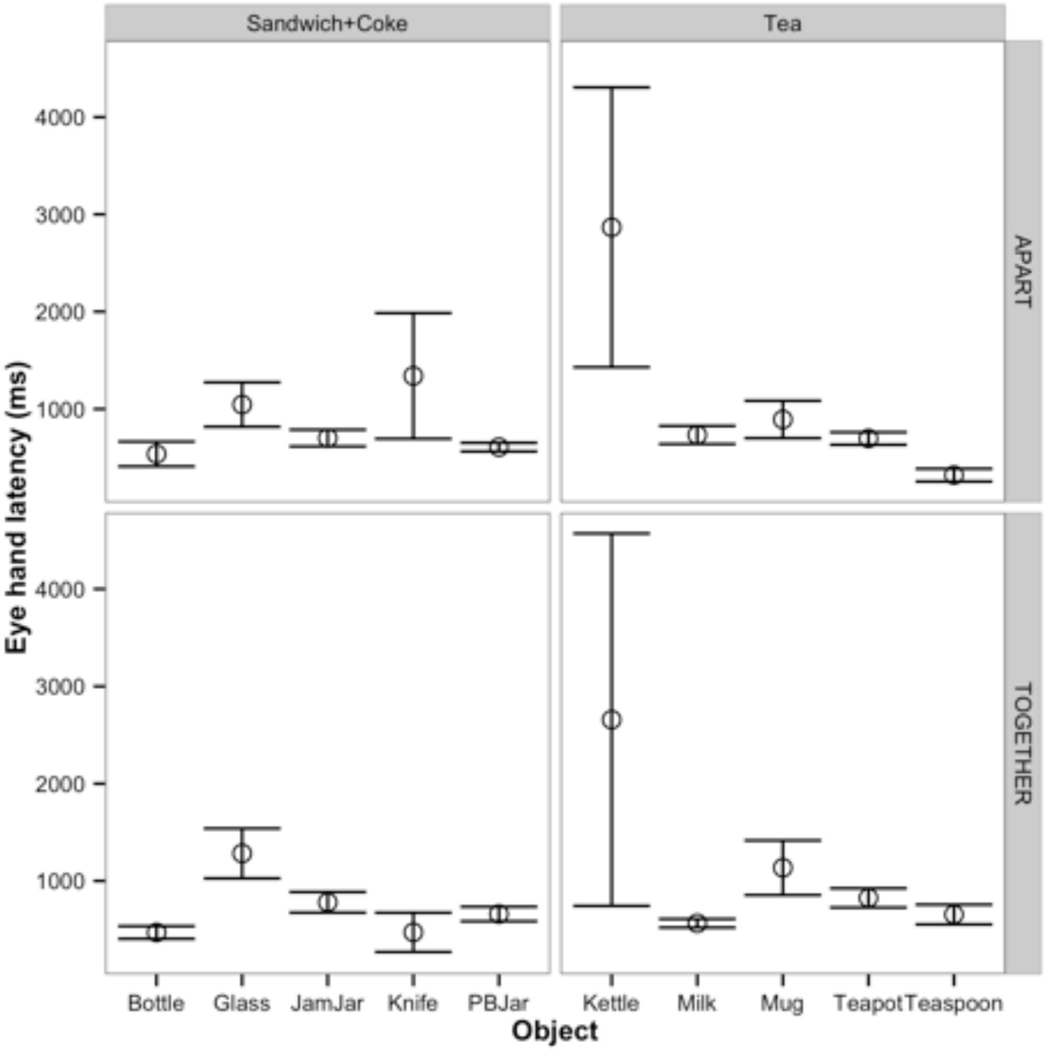


Figure 3.5 Eye-hand latencies for a subset of objects in both together and apart conditions

3.4 Discussion

We explored the link between vision and action in two complex everyday tasks: making tea and making sandwiches. Specifically we considered whether aspects of visuomotor coordination were sensitive to either the overall behavioural task or the need to move around the environment. Land and Hayhoe (2001) had previously speculated that differences found between their independently run studies were due to participants having to locomote during tea making because of the layout of the objects in the environment, compared to the stationary nature of the sandwich making task. The present study attempted to tease apart whether any differences in visual behaviour were indeed caused by the environment or in fact if the difference was caused by the tasks themselves being different. We found that task goals influenced the tendency to look-ahead of the current action (to targets of future actions), the tendency to fixate task-irrelevant objects in the room, the temporal coordination at the start of an object related act, and the tendency to use vision to guide an object put-down. The need to move around the environment influenced the extent to which people fixated task-irrelevant objects, and tended toward influencing the percentage of object put-downs that were visually guided, but did not influence other aspects of visuomotor behaviour.

The number of ORAs executed to complete each task did not vary between tea and sandwich making, or according to the two layouts of objects. Thus the two tasks were quite comparable in terms of the number of component actions across the four experimental conditions. Therefore differences in the remaining dependent variables must reflect subtle differences in visual behaviour and visuomotor coordination, rather than gross differences in task organisation.

The tendency to look-ahead to objects that are the target of future actions was generally in line with previous reports, which have suggested that around 20% of fixations can be classified as look-aheads when building models, making sandwiches or washing one's hands (Mennie et al., 2007; Hayhoe et al., 2003; Pelz & Canosa, 2001). In our study the frequency of look-ahead fixations was higher when making tea than when making sandwiches, showing that this behaviour is task sensitive. Somewhat surprisingly, whether the objects of future actions were close by (in the objects together condition) or further away (in the objects apart condition) did not influence the tendency to look-ahead. Why task differences arise here in an otherwise similar environment is not clear, but may reflect the need and opportunities to plan ahead in the task at hand: perhaps making tea permits or requires that more information is gathered prior to the execution of actions.

How frequently task-irrelevant objects were fixated was sensitive to both the task and the layout of objects. The prevalence of task-irrelevant looks was similar to that found in previous studies of tea making (around 5%, Land et al., 1999) and sandwich making (16%, Hayhoe et al., 2003). It is worth bearing in mind that there were more distractor objects in the sandwich making study conducted by Hayhoe and colleagues which might have contributed to the higher percentage of looks to task irrelevant object, because of this difference in the set up of the two environments it is not possible to interpret the direction of results as an effect of task, but the present experiment allows us to look at this.

A key aspect of this work was to consider the temporal coordination of gaze within an object related act. Characterising visuomotor coordination and the factors that influence it provides insights into the basic building blocks of behaviour. We found eye-hand latencies in our two tasks that were broadly in line with the vast majority of previous studies of visuomotor coordination in real world settings, which have found that vision tends to lead action by between 0.5 and 1 second (Land & Tatler, 2009). In the present study we demonstrated for the first time that this eye-action latency is sensitive to task demands, but not to whether the participant moves around the environment or not. While previous studies have found strong links between task demands and the spatial allocation of vision in a range of situations (Henderson, 2007; Tatler et al., 2011) we are the first to extend this to demonstrate that real world tasks influence the temporal relationship between vision and action.

Specifically, we found that eye hand latencies were shorter when making sandwiches (0.57 seconds) than when making tea (0.72 seconds) confirming in a broad sense the previous findings of Land and Hayhoe (2001). Prior to the present study it was not clear why this difference between tea and sandwiches had been observed. Land and Hayhoe (2001) speculated that these previously reported differences were likely to have arisen from the fact that participants moved around the environment in the tea-making study, but did not move around for the sandwich-making study. When moving around, this necessarily imposed a slower ‘tempo’ for the task, with longer between each object related act, and therefore greater opportunity to fixate an object longer before the hand made contact with it. Here we have shown that when the environmental setting is controlled and the need to move around is manipulated, this factor does not influence eye-hand latencies. Thus, the lead of the eye over action is not a consequence of opportunities to look at objects sooner and is a strategy employed by the visual system irrespective of how objects are distributed around the environment.

The possibility that task tempo and the need to move around to gather objects are not responsible for variation in eye-hand latencies has been raised in recent work looking at visuomotor coordination in a speed stacking task (Foerster, Carbone, Koesling & Schneider, 2011). This task proceeds with very high tempo and all objects (stacking cups) are within reach throughout the task. However, eye-hand latencies were found to be on average 0.42 seconds, which is comparable to the latencies found in Land et al.'s (1999) tea making studies.

It is important to note that the present study had a similar number of participants to the original studies, in this case 10 participants who each completed 4 trials, whereas the original sandwich making study used 11 participants who each completed only a single trial and in the original tea study only 3 participants were recorded, all for one single trial. Furthermore, the Land and Hayhoe (2001) paper compared two separate studies essentially drawing conclusions from between subjects, whereas the present study was a within subjects design thus resulting in increased power. Therefore, if the effect of movement was robust, considering both the participant numbers and the design of the study, we would have expected to have observed a difference in visual behaviour if the contributing factor was locomoting rather than task. If differences in eye-hand latencies cannot be explained by the requirement to move around the environment, they must arise from factors related to the tasks themselves. This may be considered somewhat surprising given that the two tasks are in many ways very similar: both are domestic tasks with similar constraints and the same environment. The differences between these two tasks arise primarily from the objects that are used and the actions that are carried out with these objects, raising the possibility that temporal coordination of vision and action may depend upon the specific object and our intended use of the object.

A variety of lines of evidence suggest that our visuomotor interactions with objects may vary depending upon our intended actions with that object and our expectations about the properties of that object. When approaching obstacles in a virtual environment, when and where participants directed their eyes to the object depended upon whether they intended to collide with or avoid the object (Rothkopf et al., 2007). When intending to collide with an object in this setting, participants directed their eyes to the centre of the objects, whereas when intending to avoid the object they directed their eyes to the margins of the object. Furthermore, gaze was directed to (and away from) the object sooner when the object was to be avoided, than when it was to be collided with (see also Tatler et al., 2011). Thus our intentions to act upon the object varied both the spatial allocation of gaze within the object and the temporal relationship between vision and action. Our expectations of the likely properties of objects can also influence how we interact with them. When we use a precision grip to lift an object, grip and lift forces are scaled to the expected weight of an object we are familiar with (see Flanagan, Bowman & Johansson, 2006 for a review). It is therefore possible that differences in the objects and their use within the tea making and sandwich making tasks may be the reason that eye-hand latencies differ between these two tasks.

It should be noted that while our findings are consistent with the direction of eye-hand latencies found between tea making and sandwich making in previous studies, the mean eye-hand latencies in our experiment were different from those in previous studies. We found latencies 0.57 seconds for sandwich making, which is considerably more than the 0.09 seconds reported by Hayhoe (2000). Our mean eye-hand latency was 0.72 seconds for tea making, which is more than the 0.56 reported by Land et al. (1999). It is not clear why these differences arise between our study and these previous studies. However, our eye-hand latencies in all four experimental conditions fall within the typical range of eye-hand latencies found across a broad range of activities in the real world (see Land & Tatler, 2009).

The eye hand latencies at the beginning of ORAs reveals the temporal nature of the complex interplay between the visual system and our motor behaviour, however an interesting aspect of the spatiotemporal deployment of vision during action arises when we place an object on a surface after the completion of ORA. Often this is preceded by a fixation of the empty location on the surface where the object will be placed; thus the placement is visually guided during its execution. However, sometimes these object put-downs are unguided, with no fixations made on the object of target location while it is placed on the surface. In general, more unguided object put-downs were found in our experiments (35.2% when making tea; 51.7% when making sandwiches) than in previous studies of these two tasks (5% when making tea, Land et al., 1999; 13% when making sandwiches, Hayhoe, 2000). Thus we confirm the previous finding of differences between the two tasks, but not the typical prevalence of this behaviour.

Land and Hayhoe (2001) suggested that their observed differences in the prevalence of unguided object put-downs might arise from whether or not participants had to move around an extended environment to complete the task. When moving, there may be less opportunity to encode the surroundings sufficiently to support unguided put-downs. On the other hand, information that guides hand movements may continue to be updated during orienting saccades (Cameron, Enns, Franks & Chua, 2009). When the objects were apart in our experiment there will have been more large orienting saccades between objects, as the participants move around the kitchen, and as such we might predict more opportunities for spatial encoding as therefore more unguided put-downs. Our data do not distinguish these possibilities but are in the direction of the predictions made by Land and Hayhoe (2001), we found a trend that approached significance for more unguided put-downs (59.8%) when objects were all within reach, than when the participants had to move around the kitchen (53.3%). Moreover, Land and Hayhoe's suggestion might explain why we found much greater incidence of unguided put-downs in our study than in their previous work. In our experiment, participants carried out the tasks four times in the same kitchen, providing opportunities for encoding the spatial surroundings. This familiarity with the environment may have resulted in the high prevalence of unguided put-downs that we observed.

3.4.1 Conclusion

Our findings provide new insights into the microstructure of an object related act, and thus into the coordination of vision and action in real world activities. Visuomotor coordination in time (eye-hand latencies at the start of an ORA) and space (the need to visually guide an object placement) varies between essentially similar tasks. Understanding what factors influence visuomotor coordination within an ORA allows insights into the organisation of complex behaviour: if we can characterise the properties and sensitivities of the building blocks of behaviour we can use this to better understand how complex behaviours are organised. It is clear from our findings that the microstructure of an ORA is sensitive to the requirements of the action, but not to the manner in which we move within our environment. Different tasks require interactions with different object types, the extent to which the properties of the objects used in the actions needed to complete an overall task have yet to be looked at and may be an important factor in understanding why we observe differences in visual behaviour between two very similar tasks.

Chapter 4 The influence of object properties on visual guidance during putdown

4.1 Introduction

When completing a natural task which requires action, there are several factors that may influence our general behaviour (including our visual behaviour): the environment we are conducting a task in, the task itself, the level of familiarity we have with the task and/or environment and the objects themselves. The results from Chapter 3 suggested that differences in eye-hand latencies and the visual control and monitoring of object putdowns differed depending on the task being completed. These were both similar domestic tasks completed in the same environment with the same experimental manipulations; the only things that changed were the objects used in the two tasks and the respective sub-goals. The extent to which properties of objects, such as their size, shape and functional characteristics impact on our visual behaviour during an active task is the focus of this chapter.

Being able to actively function in an environment often requires interactions with objects. In order to perform these interactions we must be able to extract information about the object, plan our action and program our related motor control. As explained by Briscoe & Schwenkler (2015) the spatial layout of the environment for example the locations of the objects, (the 'where' pathway of Milner & Goodale, 1995; 2006) can be used for motor programming of an action whilst high-level, representations of the categories of objects alongside the object’s functional and material properties (the ‘what’ pathway of Milner & Goodale 1995; 2006) are used for action planning, i.e., choosing the target to act upon or deciding what type of action is most appropriate. Object properties have been demonstrated to have considerable effect on our behaviour in the motor-planning research, both in terms of extrinsic properties of the object, such as its egocentric distance and direction and the intrinsic properties of the object (such as its size, shape and surface properties (Jeannerod, 1981, 1986; Jeannerod, 1999; Paulignan, Frak, Toni, & Jeannerod, 1997). It has been demonstrated that during reaching for and grasping for an object, the aperture of the grasping fingers increases with the size of the object to be grasped (Marteniuk, Leavitt, MacKenzie, & Athenes, 1990), similarly volume, shape and familiarity of an object has been found to influence our grasping behaviour (Gentilucci et al., 1991; Gentilucci, 2002; Goodale et al., 1994). Object properties have also been demonstrated to cause errors of perception in some cases. Van Doorn and Savelsbergh (2007) and Otto-De Haart, Carey, and Milne, (1999) used the Müller-Lyer configuration to examine perception and action and found that perceptual judgements were affected by the orientation of the arrow heads but the control of hand aperture was not, suggesting that the system is flexible enough to deal with and compensate for perceptual errors and maintain motor competency.

Properties of objects have also been shown to affect our visual behaviour. Hayhoe et al., (2004) used a simple ball catching task and demonstrated that with a little practice catchers could anticipate where the ball would bounce and direct their fixation to the future bounce point approximately 53 ms before the bounce. When the ball was covertly swapped for a bouncier, faster bouncing ball, catchers took only three practices to learn the new properties of the ball and its bounce point and again returned to making anticipatory eye movements.

Demonstrating that our visual behaviour is influenced by the properties and consequential behavioural functions of objects, and is adaptive to circumstances. From this we can conclude that the visual system employs flexibility, which allows knowledge of object properties to be updated and this high-level information used to aid performance in a motor task.

The ball catching task demonstrates the speed at which the visual system is able to take up and integrate new information about the properties of an object and then plan motor behaviour accordingly. However repeated experience in the world allows us to build up knowledge, which we can use to guide our behaviour and even our eye movements. This can be witnessed in the sandwich making study conducted by Hayhoe, Shrivastava, Mruczek, and Pelz, (2003) who noted that participants exhibited different viewing behaviours for carrying out the same action when there was variation in the properties of one of the component objects. The authors found that when peanut butter was being spread on the bread, gaze was targeted at the point on the bread where the tip of the knife would begin spreading. Ballard & Hayhoe (2009) discussed this finding and argued that participants were taking advantage of the fact that peanut butter reliably sticks to the knife and does not require constant monitoring or visual guidance. In contrast, jam/jelly is much more fluid and more precarious on the knife and thus is guided to the bread with a pursuit eye movement. The authors argue that this type of knowledge, about the properties of jam versus peanut butter, demonstrates the contribution high-level information makes in gaze allocation.

The individual properties of objects appear to influence visual behaviour before the onset of actions and during the manipulation, however little work has been done examining this relationship at the end of object manipulations in a task. We already know that during an active task fixations are made to locations where, as yet, there are no targets, but where a target is about to be set down and that the setting down of an object can either be visually guided or unguided. For example, in the tea making study 5% of object putdowns were visually unguided (Land, Mennie, & Rusted, 1999) whereas in the sandwich-making task, 16% were unguided (Hayhoe, 2000). Land and Hayhoe (2001) argued that having to locomote during the tea-making task may have imposed a slower tempo on the task and thus incurred differences in visual behaviour, such as a higher prevalence of unguided putdowns. In Chapter 3 of this thesis we demonstrated a much higher prevalence of objects being put down with no visual guidance, when making sandwiches (51.7%) than when making tea (35.2%); however we did not find support for the suggestion that movement in the tea-making task incurred more unguided putdowns.

The planning element of the initial reach and action on an object has been well studied, however the setting down an object is an element of action which requires planning, and the interaction of visual and motor behaviour to reach a goal, but as yet has received little attention. During the set down of an object there are several challenges that the visual cognitive system has to deal with, firstly the object has often changed in some way after the main manipulation, for example if the sub-goal was to pour tea from the teapot then the teapot will have reduced in fullness, thus the properties of the object have changed and therefore the visual behaviour required to complete the setting down may also change. Secondly, the set down location is unlikely to be the exact spot where the object was originally located (unless that is part of the task) therefore the area of set down may not have received previous visual attention and therefore be less familiar. Thirdly, the area of set down may contain several other objects that require vision to avoid collision. Fourthly, the setting down of an object typically requires contact of the object and a surface, the surface may not have been touched do would have not had the opportunity of any haptic feedback about it’s properties, thus the putdown relies solely on visual processing of the surface. Fifthly, the risk of misjudging some detail of the putdown, for example the height of the surface or the fragility of the object could potentially incur dropping, breaking or spilling the object, its contents or other surrounding objects. Finally, the end action of most manipulations where the object has been held in the hand, is that the object must be put down again. In a sequential task with many objects to be manipulated, the next object can usually only begin to be acted upon after the previous object has been put down (except for instances where participants use both hands to perform separate manipulations). Often we are interested in the overall tempo and the time at the start of an action between the eyes arriving and the hand arriving (the eye-hand latency). One of the factors that could potentially have a significant impact on this is the behaviour during the set down of an object, particularly the point at which the eyes leave an object and are free to fixate on a new target. Therefore visual behaviour during the putdown of an object can reveal important information about the processes occurring during the specific action in much the same way as the study of initial reaching and action planning and execution.

Object properties have clearly been demonstrated to have a significant impact on our behaviour and eye movements. In the present study we are interested in whether it is possible to exploit inherent properties of objects so that the effect on our visual treatment of objects during the setting down phase of an ORA can be examined. Using objects (drinking glasses) which have some inherent unchangeable properties (such as its shape and material) and some properties which can change such as its contents, we investigated the incidence of guided putdowns compared to unguided putdowns and examine whether there is a difference in the type of putdown made depending on the properties of the object being manipulated. We hypothesise that the place an object is to be set down is also likely to have an impact on the visual guidance during an object put down, in that cluttered put down areas will typically result in more guided putdowns than an uncluttered area since again risk of knocking other objects would likely increase the more cluttered the setting down are is. Therefore we also manipulated the putdown conditions to have both a relatively uncluttered area (4 glasses) and a cluttered area (8 glasses). Furthermore, since there are many ways to guide an object to its final resting point, we are also interested in exploring the category of unguided putdown to potentially classify at a more detailed level the exact types of visual behaviour that constitute the guiding of an object put down, and to investigate whether there is an effect of the object properties on the type of guided putdown made.

4.2 Method

4.2.1 Participants

Five female undergraduate psychology students from the University of Dundee participated in the study in return for course credits. All participants had normal or corrected-to-normal vision.

4.2.2 Materials

A laboratory in the University of Dundee, School of Psychology building was utilised for this study. The room contained one bench style worktop area where the majority of the experimental stimuli were placed, along with the computer displaying task instructions, and the workspace used to carry out the task. The set up of the glasses, monitor, jugs and fill and set down locations can be seen below in Figure 4.1. In total, 44 glasses were used, (champagne flutes, high-ball, tumblers and wine glasses) all presented in both glass and plastic versions. There were 6 glasses of all types present in both glass and plastic, except for plastic wine glasses of which there were 4. Two jugs were used, one ceramic opaque and one clear glass. Ten plastic dinner trays were used as the final set down location. Instructions were displayed on a monitor and participants were instructed to click the mouse to receive a new set of instructions for each of the 18 trials.

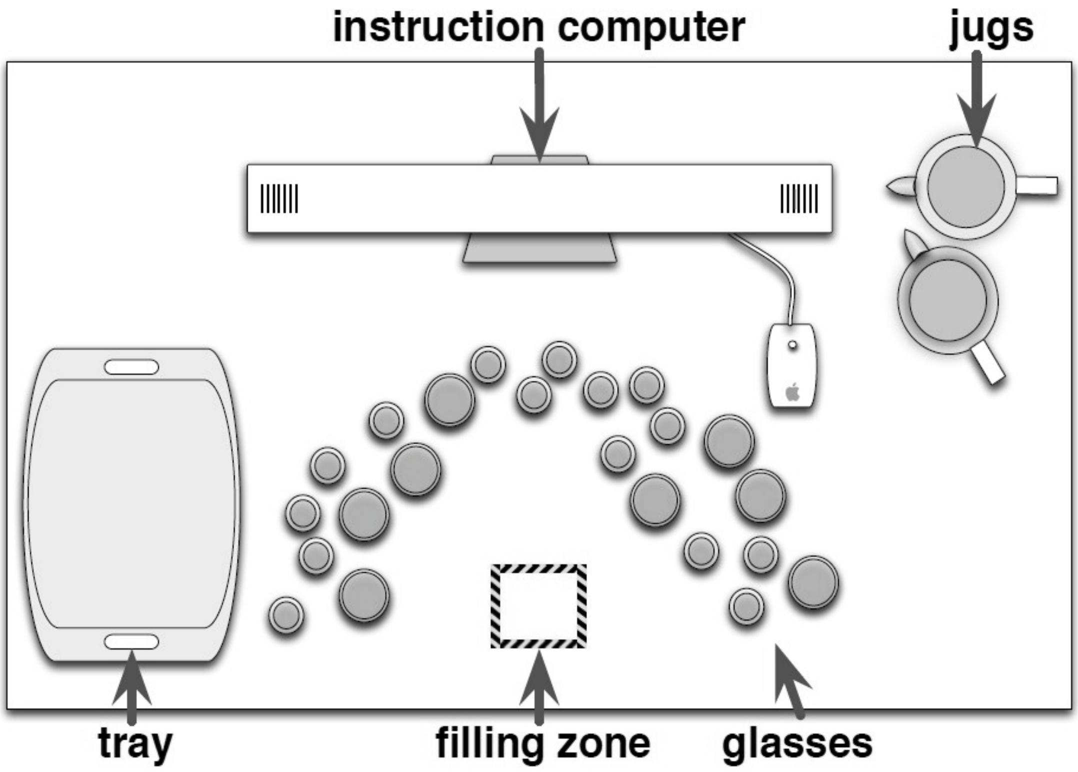


Figure 4.1 Layout and materials used for glass putdown study.

4.2.3 Eye Movement Recording

Eye movements were recorded using the same equipment and procedures as outlined in Chapter 2.

4.2.4 Design

We used a within-subjects design to manipulate four independent variables. We manipulated the properties of the glasses by using four different types, champagne flutes, tumblers, high balls and wine glasses (resulting in two tall glasses and two short) with all four glass types present in both glass and plastic. We manipulated state, by requiring participants to fill the glasses using a jug of water. The glasses were to be either left empty, filled halfway or filled all the way up. The glasses were set down empty location 1 (a filling station), then either left empty, half filled or completely filled and then picked up and set down again in location 2 (on a tray). Each tray was filled with either 4 or 8 glasses, thus the clutter of the tray was a continuous manipulation (the tray became more cluttered as more glasses were put down). The fixed effect in the LMM was continuous and ran from 1-8. Each tray contained either 4 or eight glasses before it had to be moved across the room to be set down and cleared. The dependant variable was the type of putdown made with the object in both states and the visual guidance on moving and setting down the full tray of glasses.

4.2.5 Procedure

Both verbal and written instructions (displayed on the monitor) for the task were provided at the start of the session. Participants were asked to click the mouse for each set of instructions which would inform them of which specific glass to select, they were then to set it down in the ‘filling station’, and either leave the glass empty or fill it half-full or all the way full and then transfer the glass over to the plastic tray. Each instruction screen displayed either 4 or 8 glasses to be filed and each line of instruction contained the glass type, material, fill level and jug to use for fill.

After completing all 4 or 8 instructions (glass fills and put downs) on screen participants were told to click the mouse, the monitor would then display an instruction to pick up the completed tray of glasses and transfer to a table behind them where they were to empty the glasses and set the empty glasses down on the table, before returning to the task area to click for the next set of glass filling instructions.

18 screens of glass filling instructions were displayed totalling 118 glasses used by each participant. Of those, one third were empty, one third half-full and one third full. 50% of the trays were to contain 4 glasses (uncluttered) and 50% of the trays were to contain 8 (crowded). The task in total took between 40-60 minutes and two breaks were scheduled for the participant. During the breaks, the researcher replaced the glasses in the horseshoe configuration (sticker dots were on the worktop marking position) and the jugs were re-filled.

4.2.6 Analysis

Data were discarded for an entire participant if the calibration procedure could not provide a reliable estimate of gaze position, or if any recording session was not recorded properly (e.g., due to recording failures). After these strict exclusion criteria, data from 5 participants were available for subsequent analyses (all female).

Using the same method as the previous chapter, gaze-fitted movies from each recording session were analysed manually on a frame-by-frame basis. For all eye movement related measures, we recorded gaze events rather than individual fixations. Only gazes made during the putdown of objects were recorded. Object putdowns were considered either as guided, if gaze was directed either to the object or the location the object was eventually set down on, or as unguided, where no visual guidance was used for the entire setting down process. Since there are multiple ways in which one can visually guide the putting down of an object, guided putdowns were further classified into the seven categories described in the general methods chapter (see Table 2.3 for details).

We coded the glass used during each manipulation by type (wine glass, champagne flute, tumbler and high-ball), material, and for the second putdown state (level of liquid fill; empty, half full or full) and analysed the type of putdown made both in pre-manipulation state (i.e. glass empty) and in post manipulation state (empty, half-full or full) for two putdowns, one in filling station and one on the tray. The tray was coded as either crowded (8 glasses) or uncluttered (4 glasses).

4.3 Results

4.3.1 Putdowns on the filling station

A GLMM was run using the *lme4* package in R (Bates, Mächler, Bolker, & Walker, 2015) on whether a glass was set down with (score of 1) or without (score of 0) visual guidance. The maximal converging model included glass type and material as fixed effects, with participant and trial number as random effects. The full output of the model is shown in Table 4.1. In sum, participants were most likely to set down champagne flutes with visual guidance. While this probability was not significantly different to the probability of guiding a high-ball glass, it was higher than the probability of guiding both the tumbler and wine glass. There was no significant effect of whether a glass was made of glass, or a plastic equivalent.

Table 4.1. GLM results for putdown on filling station

DV	Test level	Comparison level	Beta	SE	z	p
Glass type	High ball	Champagne flute	-0.37	0.3	-1.2	0.23
	Tumbler	Champagne flute	-0.9	0.3	-3.04	<.01**
	Wine	Champagne flute	-0.62	0.3	-2.08	<.05*
Material	Plastic	Glass	-0.28	0.22	-1.29	0.2

4.3.2 Putdowns on the tray

A GLMM was run on whether a guided put-down was used when setting the glass on the tray. This model was the same as the previous model, except that we also included the fill-level of the glass (empty, half-full or full), and the ordinal number of the glass being set down on the tray, as this would increase within a trial, making the tray more crowded. The full effects are shown in Table 4.2. When setting down the glass on the tray, the effect of glass-type shown when placing glasses in the filling station was not found. Conversely, the material that the glass was made of significantly influenced the probability of visual guidance, with glasses made of glass being more likely to be guided than plastic ones. While there was no difference in guidance between empty and half-full glasses, there was a significant difference between empty and full glasses, with the latter being more likely to be guided. Finally, the strongest contributor to guidance was observed in the glass number, with participants being more likely to guide glasses as the tray became fuller.

Table 4.2 GLM results for putdown on tray

DV	Test level	Comparison level	Beta	SE	z	p
Glass type						
	High ball	Champagne flute	0.03	0.75	0.04	0.97
	Tumbler	Champagne flute	-0.69	0.63	-1.1	0.27
	Wine glass	Champagne flute	-0.09	0.65	-0.14	0.89
Material						
	Plastic	Glass	-2.01	0.53	-3.8	<.001***
Fullness						
	Half	Empty	1.0	0.57	1.76	0.08.
	Full	Empty	1.75	0.61	2.86	<.01**
Glass number						
	.	.	0.81	0.2	4.09	<.001***

4.3.3 Types of guided putdowns on the tray

The level of visual guidance during a guided putdown can vary. The exact breakdown of types of visual guidance typically deployed during the set down of a n object are described in detail in Table 2.3, Chapter 2. Broadly, guidance can be grouped into two categories: continuous visual guidance (score of 1) and non-continuous (score of 0). For all guided putdowns the type of guidance (continuous and non-continuous) was analysed the results reveal the same pattern as the previous two sections, continuous visual guidance was most likely to be applied if the glass type was a champagne flute, if the fill level of liquid was full or if the set down surface was crowded. The full effects are reported in Table 4.3.

To examine whether the types of guidance during glass putdowns changed across the duration of the experiment, we compared a model that did, and did not include tray number. We found that there was no significant difference between these models (p = 0.29), suggesting that there were no changes across the experiment. Similarly, the probability of unguided putdowns did not change across the duration of the task, (p = 0.54).

4.3.4 Putdowns of full trays after each trial

For each participant all of the 18 trays, regardless of whether they contained 4 or 8 glasses dedicated the same pattern of visual guidance during the setting down of the full tray. All trays (100%) were continuously guided during the putting downs phase.

Table 4.3 GLM results for putdown on tray – guided and part-guided

DV	Test level	Comparison level	Beta	SE	z	p
Glass type						
	High ball	Champagne flute	-0.91	0.45	-2.03	<.05*
	Tumbler	Champagne flute	-1.14	0.44	-2.57	<.05*
	Wine glass	Champagne flute	-0.52	0.41	-1.26	0.21
Material						
	Plastic	Glass	0.50	0.33	1.51	0.13
Fullness						
	Half	Empty	0.56	0.43	1.31	0.19
	Full	Empty	1.39	0.43	3.24	<.001***
Glass number						
	.	.	0.25	0.09	2.77	<.01**

4.4 Discussion

Prior knowledge, current visual information and sensorimotor feedback about the properties of objects are combined to evaluate whether continuous visual guidance is required during the setting down of an object, this evaluation is flexible and responsive to changes in the properties of objects. We manipulated glass type and material, liquid fill level and crowding of the set down surface in order to examine the effect of object properties on visual guidance during setting down. Putdowns are more likely to be visually guided depending on the glass type, material and fill level, but this changes depending on the context. If glasses are empty, then the glass type is the most important factor that will determine if the putdown is guided or not but when the glasses contain liquid this along with the effect of glass type contributes to the decision to perform the put down with visual guidance or not. The number of glasses that were already on the put-down surface accounts for the most amount of variance in determining whether the glass putdown will be guided or not, as the setting down surface becomes more crowded the likelihood of visual guidance during put-down increases.

Prior knowledge of the properties of objects impacts the visual guidance of the setting down action from the onset of a task. We found no change across task in terms of the types of putdowns made; if making unguided putdowns was a product of learning then we would have expected at least the first tray to have been performed with guided putdowns until the participants had learned the properties of each object during the set down phase. It has previously been demonstrated that two distinct gaze behaviours (visual guiding of a target movement or unguided target movement) are elicited as a feature of the learning process, with unguided behaviours emerging as learning of the task is acquired (Säfström, Johansson, & Flanagan, 2014). Land, Mennie, and Rusted (1999) and Hayhoe (2000) demonstrate that visually unguided behaviours also occur in real life everyday well learned tasks, with a significant proportion of specific actions receiving no visual guidance, typically these are instances where an object is put down on a surface. We find that there is no familiarisation process required to learn the properties of objects and whether they would benefit form visual guidance during putdowns, indicating that we are able to rely on internal models of the physical properties of objects and the relationships between objects in a scene (Flanagan & Wing, 1997; Johansson, 1996). However, that is not to say that we only use prior knowledge, it is more likely that we combine our prior knowledge with the present visual information others have suggested that the combination may be in the form of an optimal Bayesian integration strategy, as is demonstrated in much of the sensorimotor control literature (Körding & Wolpert, 2004; Tassinari, Hudson, & Landy, 2006). This ability to combine prior knowledge and current visual information and update and adapt to new information is crucial for living in a dynamic world where the properties of objects can change (Diaz, Cooper, Rothkopf, & Hayhoe, 2013).

Being able to use prior knowledge of the properties of objects and how they interact with our environments is crucial for the visual system to be able to make predictions and guide fixation accordingly. When setting down an empty glass in an empty location the decision to use visual guidance or not seems to be based on glass type (style), with the tallest glasses, champagne flutes and high ball glasses being more likely to be guided than wine glasses or tumblers. From this result alone, it is not clear whether this might arise from the height, weight, or material of these glass types. If it was either weight or material then the difference should have been between glass and plastic (since alongside glass being breakable, it is also heavier), however this was not the case and there were no differences between the visual guidance of glass and plastic glasses during putdowns, when the glass was empty. From our results it seems more likely that the deciding factor for champagne flutes and high-ball glasses to receive more visual guidance during the set down was due to height. We know from previous literature that visual behaviour (fixation locations) changes with incremental increases in object height for real and computer generated objects, both when making perceptual judgements and grasping objects (Desanghere & Marotta, 2011), thus height influences visual behaviour at the start of an action. If we consider this finding in conjunction with the results demonstrated by Cinelli, Patla, & Allard (2009) showing that threat to stability (in their case when locomoting through oscillating doors) concentrates fixations on specific task relevant features and elicits more “online” control to directly guide behaviour, we can begin to appreciate that taller glasses pose more risk in terms of their stability when being set down and as such require more visual guidance, hence the increase in visually guided put downs for these glasses.

Having flexibility to respond to current visual information ensures that as properties change, visual guidance can change accordingly too. Our results demonstrate this flexibility with a range of visual guidance techniques utilised depending on the current properties of the object and setting down environment. The flexibility of behaviour exhibited by the visual system has previously been well established in terms of information use and visuomotor guidance (Roach, Heron, & McGraw, 2006; Sims, Jacobs, & Knill, 2011; van Beers, Sittig, & Gon, 1999) and in the present study we found that as the property of the glasses changed so too did the associated visual behaviour. After the initial put down, glasses had to be either left empty, half filled or filled and then picked up and set down on a tray, trays were to contain either four or eight glasses. In this setting down scenario, the type of glass (height) no longer mattered; instead the important property of the glass was the extent to which it was filled and the material it was made from. Full glasses, and those made of glass were more likely to be visually guided during put down than half-full, empty, or plastic. The change in property of the glasses, in this case the content of liquid, affected the level of visual guidance during a put down, thus here we clearly have evidence that the coordination system in control of vision and action is flexible and can adapt to varying properties of objects. This result supports the findings of Sims, Jacobs, & Knill (2011) who designed a task which imposed competing demands on the visual system using a virtual workspace environment. The task was a block-sorting task, which required vision to be used both for information acquisition and on-line guidance of a motor act. To examine visual information acquisition, blocks were rotated either 45° clockwise or counter-clock wise with the aspect ratio manipulated in order to increase the difficulty of perceptual judgement. In order to examine vision during the guidance of a motor act, the size of the bins the blocks had to be placed in varied, the authors hypothesized that smaller bins would require more visual guidance to ensure accuracy. The authors found gaze to be adaptive, when the aspect ratio made the task more difficult, the block was fixated longer compared with the two easier conditions, however, less time was spent fixating the block if the subsequent bin for placement was smaller. Sims et al. (2011) argue that participants adaptively adjusted fixation allocations and durations based on the difficulty of both the task in hand and the up-coming one accordingly depending on individual varying task demands.

Further support for adaptive visual behaviour based on task demands is demonstrated in our finding that the level of clutter of the setting down surface was a contributed to visual guidance during the setting down phase; visual guidance of the put down was more likely as the trays became fuller. As found by Sims et al. (2011) once again, as an element of the task changed, the visual system responded adaptively. As the set down area became more crowded so too did the risk of making contact with another glass during the set down procedure, our results demonstrated that putdowns were more likely to be visually guided as the tray got fuller, indicating that we are able to adapt the level of visual guidance depending on the circumstances of the task and that our visual behaviour changes as the demands of the task change.

Visual guidance during putdowns appears to respond to changing properties of objects and their surroundings on a needs basis. In the first instance, when neither liquid or surface crowding were features of the task, visual guidance was influenced by the height of the glass, then the second putdown of glasses contacting liquid revealed that full glasses were visually guided regardless of their height but ultimately regardless of glass type (height) or the liquid it contained (or not) if the set-down surface was crowded then the glass was likely to be visually guided as it was set down. This variation in visual guidance depending on the changing properties of objects and their surroundings points to a flexible visual system, adaptable to new circumstances.

Just as the level of visual guidance dedicated to objects during a put down is adaptable to changes in task, properties of objects and/or the environment, we are also capable of differentiating between instances that require continuous visual guidance and those do not. It was identified during the hand coding of the data, that guidance of put downs can take several forms, for example, the participant could look only to the setting down location for the duration of the action, or to only the object, or to a combination of both, guidance such as this was considered as continuous. It was also noted that some visual guidance was not as continuous and resembled something more akin to guidance checks, for example participants would often look at the object until the final few frames then direct gaze to another object or location until the putdown was complete, or would be fixated on another location during most of the putdown phase only to fixate on the object or put down location for the final few frames, guidance of this form was still considered as guidance but not continuous. Comparing the effects glass type, fill level and surface crowdedness had on the likelihood of these two types on guidance revealed that champagne flutes were the most likely to be continuously guided as were glasses which had been completely filled and similarly as the tray surface got fuller it was more likely that putdowns would be visually guided continuously. After moving a full tray (of either 4 or 8 glasses) to a table located across the room (requiring a 180° turn) all tray putdowns were continuously visually guided regardless of the number of glasses, glasses type, material or fill level.

The way we visually behave when setting down an object after a manipulation has received little attention in the study of vision and eye movements, yet at this juncture in a task the visual behaviour performed carries over consequences for the next action in a sequential task, particularly in terms of at which point the eyes are free to fixate the next object to be manipulated. Following on from initial observations in natural tasks that a significant proportion of these putdown actions are performed without visual guidance (Land & Hayhoe, 2001) we attempted to examine what it is about certain objects that may make them the subject of guidance during putdown or not. We manipulated the properties of glasses and the crowdedness of set down areas in order to try to unpack the type object properties which influence and determine the level of visual guidance used during put downs and found that the visual system is flexible in its response to changes in object properties, adapting to new circumstances and adjusting the level of guidance accordingly. Initially the height of the object and appears to determine whether the object will be guided or not, with tall glasses most likely to be visually guided during the putdown phase, however this is only when the glass is empty. As soon as the glass is full it is this factor along with material rather than height, which demands guidance during a putdown, finally the visual system is more likely to guide putdowns as the area for setting down becomes more crowded. The findings suggest that we utilise prior knowledge regarding the properties of objects and the way they behave in the world to inform our visual behaviour and prioritise the objects with properties that demand a higher level of visual guidance, and that this is flexible and adapts to changes in properties accordingly. What we do not know from the present study is whether the next step in sequential actions also influences the likelihood of an object being put down with or without visual guidance. It may be that the properties of the object to be used in the next step demand more processing and elicit a longer fixation thus increasing the likelihood of cutting short the visual guidance of the putdown of the present object. Alternatively it may be to do with having to remember instructions for a task or steps in sequential tasks, for example in the present study it is plausible that a participant may have read two lines of instruction at once and upon putting down one glass may have cut short the visual guidance in order to more quickly fixate the next object, this may reflect some cognitive offloading of the instruction that was being held in memory or it may be to do with the properties of the object used in the next step. It is not possible to separate these potential influencing factors related to the next step in the task from the present study but this would be an interesting issue to tease apart with further study.

The properties of objects influence the way we visually behave when performing actions with them and the present study has demonstrated that the area an object is about to be put down in is also a factor in determining our level of visual guidance, in that, we are more likely to visually guide the putdown of a glass in a cluttered area. If we consider the set down area as part of the environment of the room, it may be that other factors about the environment where a task takes place may impact our visual behaviour. More specifically, prior knowledge about an environment may change the way we visually behave during the completion of an active task.

Chapter 5 Comparing eye movements during tea making in both novel and familiar natural environments

5.1 Introduction

The previous chapter demonstrated that the specific properties of objects influence our visual behaviour when handling them. More guided putdowns were made for objects considered as less sturdy and likely to pose more risk of breaking/spilling (for example full glasses or tall empty glasses). The differences in visual behaviour for ‘risky’ objects was apparent from first handling, suggesting that participants were using prior knowledge of these common objects rather than learning the properties during the task. Prior knowledge of objects influences the way we visually behave when performing an active task, however, objects are not the only thing that we can have prior knowledge of. Most of the environments dealt with on a day-to-day basis are typically familiar, for example one’s own house, route to work and office, however familiarity of environments in the sense of prior knowledge built up over extended periods of time is not easily replicated in a laboratory since the environments that individuals are familiar with tend to be highly specific to the individuals themselves. As a result, most work examining the effect of familiarity concentrates on familiar tasks and is often conducted in unfamiliar environments rather than familiar environments.

Expertise can be essentially be regarded as the result of considerable familiarity with a task. In general it takes a significant amount of time to become an expert at something and when one does, the associated knowledge tends to be quite extensive and specific to the domain (Chiesi, Spilich, & Voss, 1979; Glaser, 1987). Without becoming embroiled in the debate regarding the definition of expertise, it is useful to briefly consider that expertise can be defined in a number of ways. Hoffman (1996) asserts that at a cognitive level, expertise can be defined in terms of its development, expert's knowledge structures and reasoning processes. The development of expertise refers to the level of skill/knowledge accumulation and several authors (for example, Spiro et al, 1989; Adelson, 1984; Gaeth, 1980) argue that during this accumulation there are level-like qualitative shifts that develop as expertise develops. The authors argue that during the development of these level type stages of expertise, it is rare for a level to be skipped, regressions or failures (not as a result of failure to practice) are rare, experts can anticipate the errors that will typically be made by a trainee depending on their skill level and that in time, with practice a skill becomes much more automated. The knowledge structures that appear to characterise expertise refer to the ability of experts to draw more complex conceptual distinctions than novices (Murphy & Wright, 1984). Hoffman (1996) also defined expertise as including advanced reasoning skill, with experts tending to spend proportionately more time at the start of a problem solving task forming a conceptual understanding of the problem and tend to demonstrate more advanced perceptual skills than novices do.

The notion of expertise affording enhanced perceptual skills has been demonstrated by several studies, with findings from comparisons of reaction times and recognition rates for novice and expert radiologists (Myles-Worsley et al., 1988) dermatologists (Norman, 1989) mammographers (Nodine, Kundel, Lauver, & Toto, 1996) and light microscopists (Krupinski et al., 2006) all suggesting that experts are faster at allocating their attention to and more accurate at recognising areas of interest which are typical of each diagnostic category. Stainer, Anderson and Denniss (2015) also demonstrated that experienced optometrists were able to fixate areas of abnormality significantly faster than novices and that experienced optometrists made fewer inspections of regions of diagnostic interest that might reveal signs of disease (the macula, optic nerve head and vascular arcades) when determining retinal health from photographs of the fundus. These results suggest that experience appears to allow optometrists to have an advantage in terms of reducing the areas to process which is simpler and faster than having to frequently re-inspect areas of interest.

Expertise also appears to have an effect on our ability to anticipate subsequent events or future areas of interest, many empirical examples of this come from sports, such as cricket (Land & McLeod, 2000) baseball (Regan, 1992; Watts, 1991) table tennis (Ripoll et al, 1987) and squash (Hayhoe, McKinney, Chajka, & Pelz, 2012). Similarly the notion of expertise affording the ability to direct gaze more accurately to areas of interest can be seen in studies examining driving. Underwood, Chapman, Bowden, and Crundall, (2002) demonstrated differences between novice versus experienced drivers, with experienced drivers conducting more visual search on challenging sections of dual carriageway road. The authors argue that novice drivers are unable to predict what is likely to occur on these demanding roads.

Vision and action are temporally coordinated thus it is important to examine not only where people look but also when things are looked at and the changes in this coordination caused by expertise. Studies examining eye movements during the learning of a novel task, for example Sailer et al. (2005) demonstrate that experience with task and knowledge accumulated during a training period (in this case 20 minutes familiarization) directly affects subsequent visual behaviour. As discussed in the introduction chapter, Foerster et al. (2011) used a bi-manual, high speed cup-stacking task for 14 consecutive days (for 45 minutes per practice session) and found that there despite the fact that where participants looked didn't change across days, eye-hand latencies did change, becoming shorter across days.

Clearly then prior knowledge and experience of a task have an influence on eye movements and even appear to afford some advantages in terms of gaze strategies and even speed of task completion (caused by reduced eye-hand latencies) however whether the same holds true for having expert knowledge regarding an environment during the completion of an active task remains to be seen. We know from work where real world scenes are viewed repeatedly that participants are faster to find targets. Brockmole & Henderson (2006) found that even after only 10 presentations of a scene, the target was fixated in less than two eye movements. Even after the image was mirror reversed participants corrected the initial fixation to the old location faster for images that had previously been repeatedly viewed. In the present study, we are interested in the effect of being experienced with an environment on visual behaviour during the completion of a natural task. In order to examine the task in both a truly novel and a truly familiar environment, the data were collected in participants' own homes (of which all had been living in for more than one year at the time of collection). Participants were paired up and performed the well-learned task of making tea in both their own (familiar) kitchen and an entirely novel kitchen (their partner's) whilst eye movements were recorded in order to reveal any differences in visual behaviour. Considering the literature on the effect of task expertise and prior knowledge on eye movements and as the previous chapter demonstrated an effect of object properties on visual behaviour, we would expect that there would be less time spent searching, reduced eye hand latencies, and less looks to irrelevant objects) for familiar environments than novel is expected.

5.2 Method

5.2.1 Participants

Six individuals (1 male) from the University of Dundee participated in the study on a voluntary basis, three of whom were undergraduate students, two postgraduate and one postdoctoral researcher. All participants had normal or corrected-to-normal vision.

5.2.2 Materials

The actual home kitchens belonging to the participants were used in this study. In order to keep the environment as naturalistic as possible, no restrictions on layout or content were in place. Before testing sessions, supplies for the tea-making task were checked and topped up if necessary. The target objects used for tea making included a mug, kettle, sink, teabags, spoon and milk, non-target objects varied in amount from kitchen to kitchen but were all items one would expect in a kitchen.

5.2.3 Eye Movement Recording

The same equipment and procedure were used as outlined in Chapter 2.

5.2.4 Design

A within-subjects design was used in order to manipulate the level of familiarity with the kitchen. Kitchens were either familiar (participants' own) or entirely novel (experimental partner's kitchen).

5.2.5 Procedure

Six participants were partnered up into three couples. Each of the couples had never been to their partner's home and had themselves lived in their home for at least one year prior to the experiment. Each member of each couple made a cup of tea both in their own home and one in their partner's home on the same day, the order was counterbalanced within couples so that one participant would complete the task in their own kitchen first then their partner would complete it in the same novel kitchen and vice versa. Participants were instructed to make themselves a cup of tea; however they normally would with no guidance as to where target objects were.

5.2.6 Analysis

Using the same method as the previous chapter, gaze-fitted movies from each recording session were analysed manually on a frame-by-frame basis. For all eye movement related measures, we recorded gaze events rather than individual fixations. Several measures of eye movements, (eye-hand latencies at both the start and end of ORAs, Change Of State beginning and end, and type of put downs made) and general behaviour (such as overall task completion time and errors) were made. Data were analysed using Linear Mixed-Effect Modelling. In this Experiment, it was important to remove any variance that might be due to the differences between the different kitchen environments, such as some objects being further away and therefore requiring locomotion. As such, we included whether the kitchen was the participant's, or their partners (our principal variable of interest) as a fixed factor, but also included the kitchen that the tea was made in as a random effect. As this factor was only repeated across the pairs, we had to simplify our mixed-effect structure to include modelling across intercepts only. The full model structure is described for each analysis.

5.3 Results

5.3.1 Overall task completion time

A LMM was fit to the task completion time data. Whether the kitchen was the participant's or their partner's was included as a fixed effect the model, and which kitchen the tea was made in (i.e. the kitchen owner), and subject were included as a random effects. There was a significant effect of familiarity level on task completion time, with tasks being completed faster in participant's own homes (*β* =87.44, *SE* = 25.41, *t* = -3.441, *p* = 0.007; Figure 5.1).

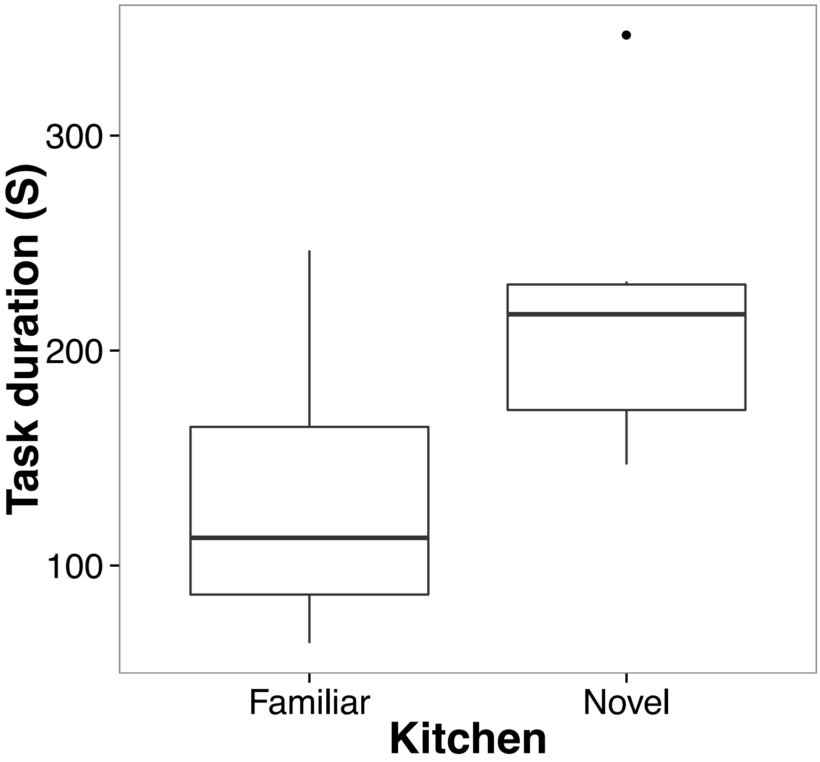


Figure 5.1 Median boxplot of task duration in familiar and novel kitchen.

To further examine this effect, we looked at the number of object manipulations that participant made in their own kitchen or their partner's kitchen. Participants made on average far fewer manipulations of objects in their own kitchen (*M* = 13.5) than in their partner's kitchen (*M* = 19.7). A LMM was used to examine this effect, and there was a significant difference between the number of manipulations of objects (*β* = 6.167, *SE* = 0.727, *t* = 8.488, *p* <.001; Figure 5.2).

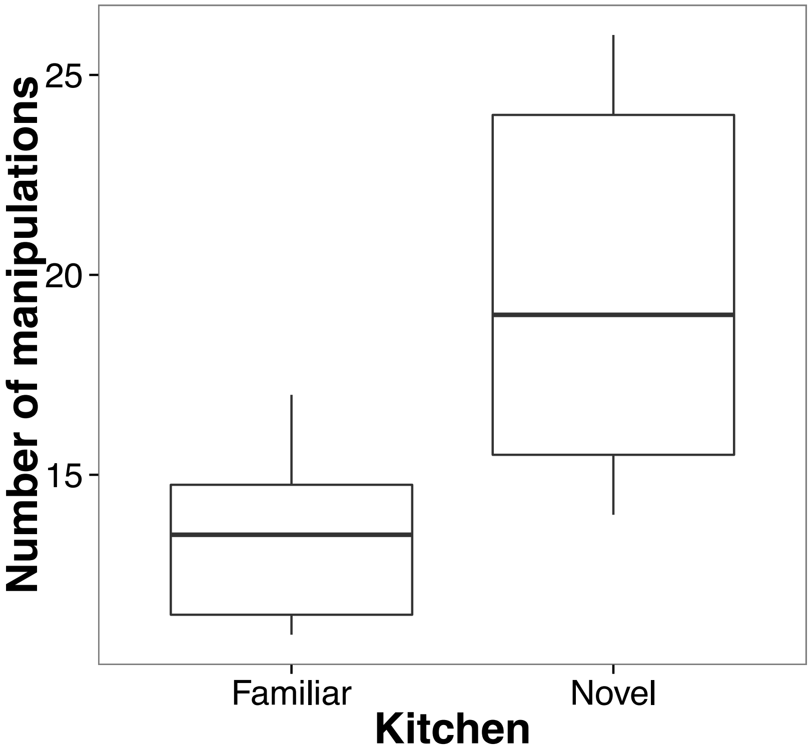


Figure 5.2 Median boxplot of number of manipulations made to complete task in familiar and novel kitchens.

5.3.2 Time spent on each element of behaviour

Despite the task being the same for both novel and familiar environment participants were significantly faster when making tea in their own kitchen than in their partners. The core number of sub-goals that must be completed were the same for both environments so it poses the question of what accounts for the time loss? Initially to get an idea of the difference a familiar or novel environment has on the way we spend our time during the task we grouped behaviour into four categories, two of which being visual behaviours that do not involve action, *look-ahead fixations* and *exploration,* and two of which are directly attached to action, *touch (all interactions which involve action/object manipulation)* and *eye-hand latency.* By consolidating behaviours into these categories we can account for changes in time spent on each behaviour type for the novel and familiar environment. We are also interested in the specific elements of visual behaviour that may be different depending on whether the environment is familiar or novel and so later consider visual behaviours, for example eye-hand latencies at an individual level.

There are two ways that we can consider time in the task – absolute time and proportional time. Absolute time is a useful measure as it provides an accurate representation of the time that participants spent on particular actions, or the temporal relationship between action elements (e.g., the time between looking at an object and touching it). However, to understand whether behaviour changes across conditions it is also important to consider changes *relative to the task completion time*. The reason that this is important is that it can tell us whether all elements of the task speed up in the same way – in other words is behaviour entirely the same, just faster, or does the visual behaviour change in different ways across tasks. By representing visual behaviour in absolute time, and *relative* to task completion time (i.e. the proportion of time) we can consider behaviour in both of these ways.

5.3.3 Absolute time spent on each element

LMM's were fit to the data where we examined the absolute time spent on each of the four elements of the task (EHL, look-ahead fixations, exploration and touch). We allowed the model to vary across subject and kitchen environment. Participants making tea in a novel kitchen spent significantly more time in total in the period between fixating and acting on an object (EHL's), just over 9 seconds on average ($\beta = 9.213$, $SE = 2.095$, $t = 4.398$, $p < .001$). The absolute amount of time spent looking at task relevant objects that would be acted on later was not different in familiar and novel kitchens (*p* = .759). However, the amount of time spent exploring the environment (*β* = 42.19, *SE* = 14.7, *t* = 2.869, *p* < .016) and the amount of time touching objects (*β* = 18.268, *SE* = 2.523, *t* = 7.242, *p* < .001) was significantly shorter in a participant's own kitchen.

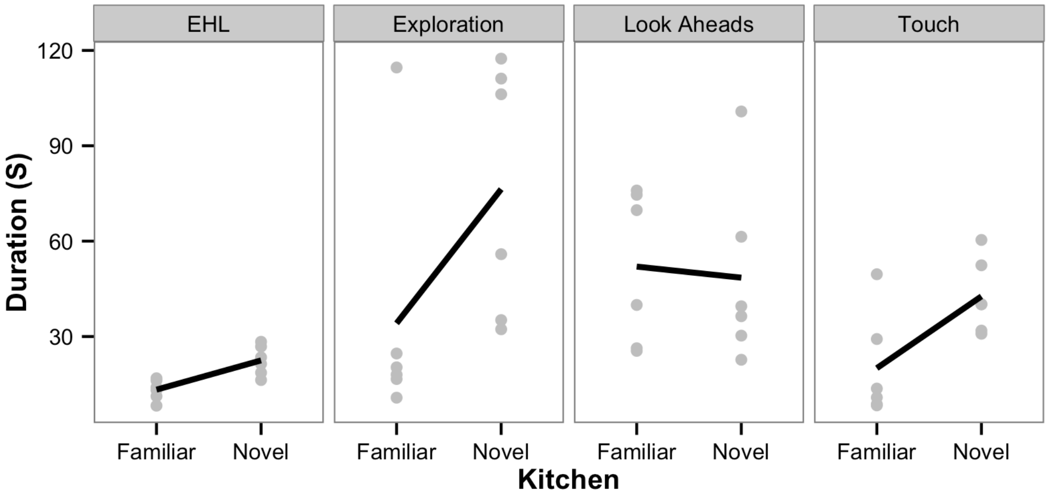


Figure 5.3 Total amount of time (seconds) spent on the four parts of the task in familiar and novel kitchens

5.3.4 Proportion of time spent on each element

The same analysis was used to examine the relative proportion of the trial spent on these acts. Given that participants were much faster at making tea in their own kitchen, this can give us a measure of performance in relation to the entire task completion time (Figure 5.4). Exploration time made up a significantly longer proportion of the trial in novel kitchens (*β* = 0.139, *SE* = 0.054, *t* = 2.574, *p* < .027), whereas the proportion of the trial making look-ahead fixations was longer in the familiar kitchen (*β* = -0.206, *SE* = 0.038, *t* = -5.371, *p* < .001). In relative trial time, there was no significant difference (*p* > .05) in the proportion of time spent on eye-hand latencies, and touching objects (although there was a slight trend for shorter amount of time touching objects in the familiar kitchen).

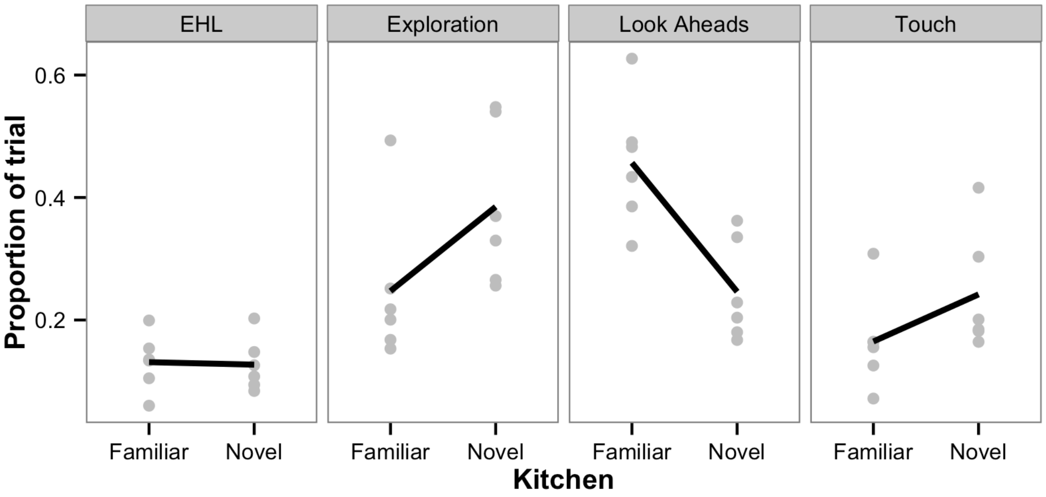


Figure 5.4 Proportion of trial spent on each of the four parts of the task.

5.3.5 Looks to task relevant and task irrelevant objects

The proportion of time spent looking at task relevant objects was significantly higher in familiar kitchens (β = -24.794, *SE* = 5.501, *t* = -4.507, *p* < .001), with participants allocating around 25% more looks to task irrelevant objects in the novel kitchen.

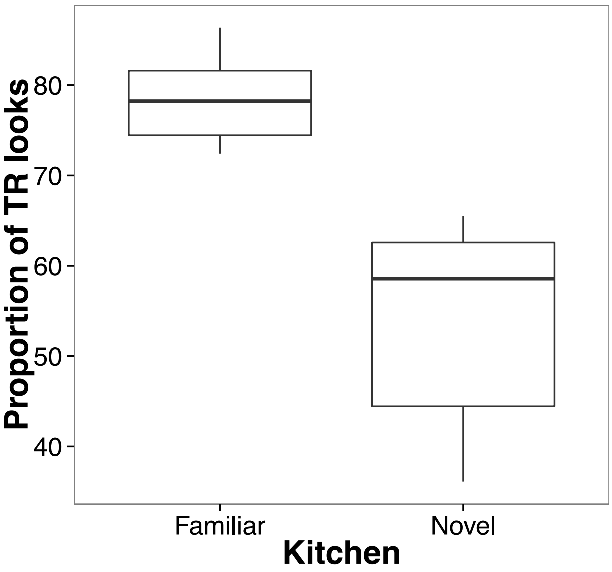


Figure 5.5 Proportion of looks to task relevant and irrelevant objects my novel and familiar kitchens

5.3.6 Eye hand latencies (beginning manipulations)

The median eye hand latency across all data was 724.1ms (Figure 5.6), on average. The eyes led the hand in the majority of cases, with only 3.54 % of the cases being where the hand led the eye (negative eye-hand latencies).

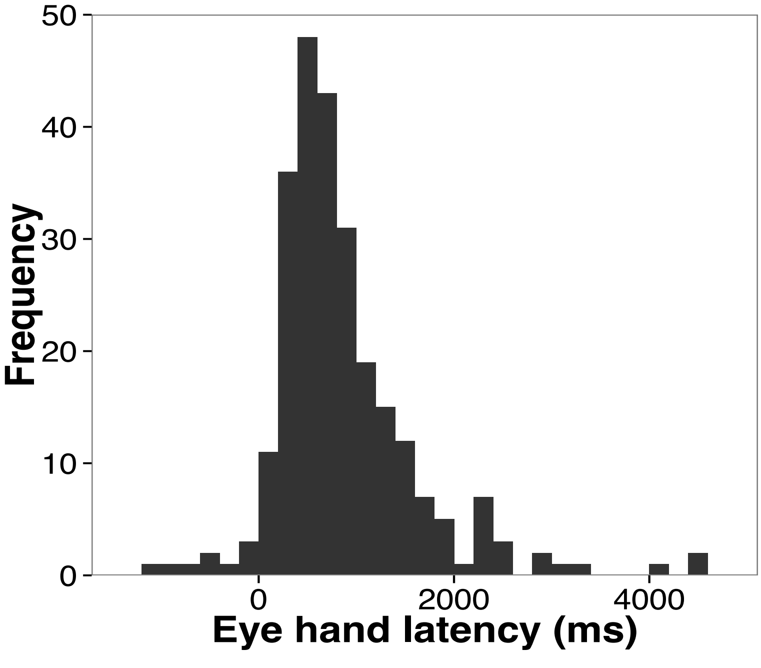


Figure 5.6 Histogram of eye hand latencies in the task.

Of the cups of tea made in participants' own kitchens, a slightly higher proportion of eye hand latencies were negative (4.08%) than when the tea making was in the partner's kitchen (3.2%)(Figure 5.7). The median eye hand latency of all of the data collected in participant's own kitchens was 637.9 ms, with a median eye hand latency of 724.1 ms in the partner's kitchen.

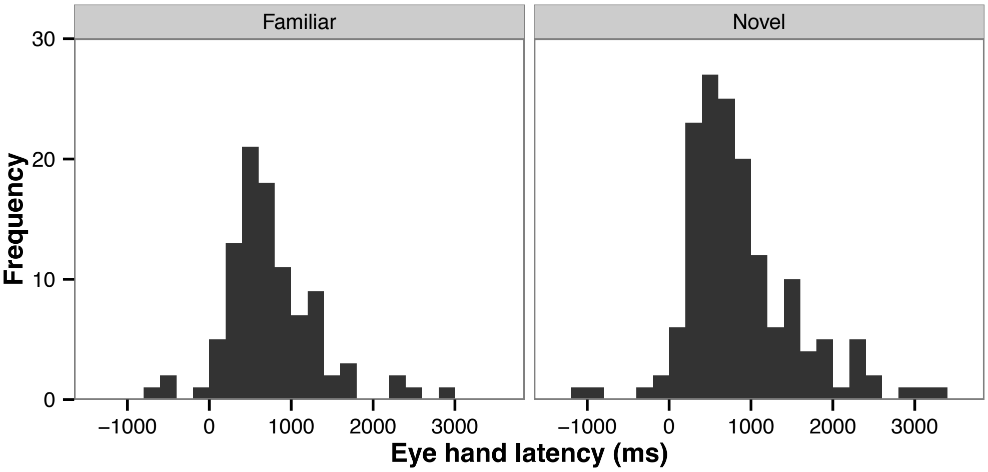


Figure 5.7 Histograms of eye hand latencies in other and own kitchen.

To examine whether eye hand latencies differed in the participant's own, or their partner's kitchen, an LMM was calculated including the type of kitchen (own or partner) and the owner of the kitchen as a fixed effect, with participant, object and the person who owned the kitchen as random effects. The kitchen owner was included to try to account for any differences that were due to kitchen layout. Eye-hand latencies were significantly shorter in participants' own kitchen than their partner's kitchen (β = -184.6, *SE* = 89.50, *t* = -2.063, *p* = .039).

5.3.7 Eye to change-of-state latencies

An LMM was fit to the eye to change-of-state latencies and there was a significant effect (β = -372.8, *SE* = 131.2, *t* = -2.841, *p* = .005). The median latencies in participant's own kitchen were 1034.5ms, increasing to 1413.8ms for cups of tea made in the novel kitchen.

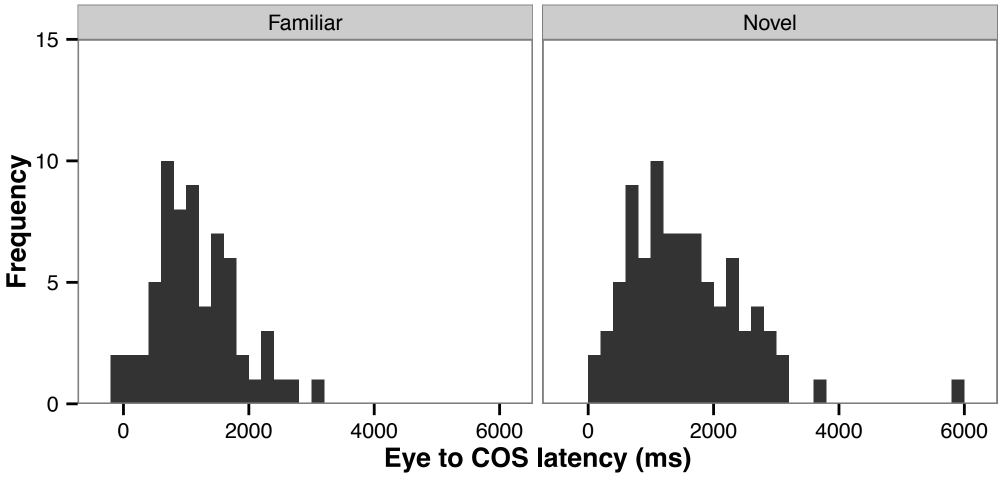


Figure 5.8 Latency of eye arrival to change of state of object

5.3.8 Hand to change-of-state

A LMM was fit to the hand to change-of-state latencies, but there was no significant effect (*p* = .056). The median latencies in participant's own kitchen were 379.3 ms, increasing to 517.2 ms for cups of tea made in the novel kitchen.

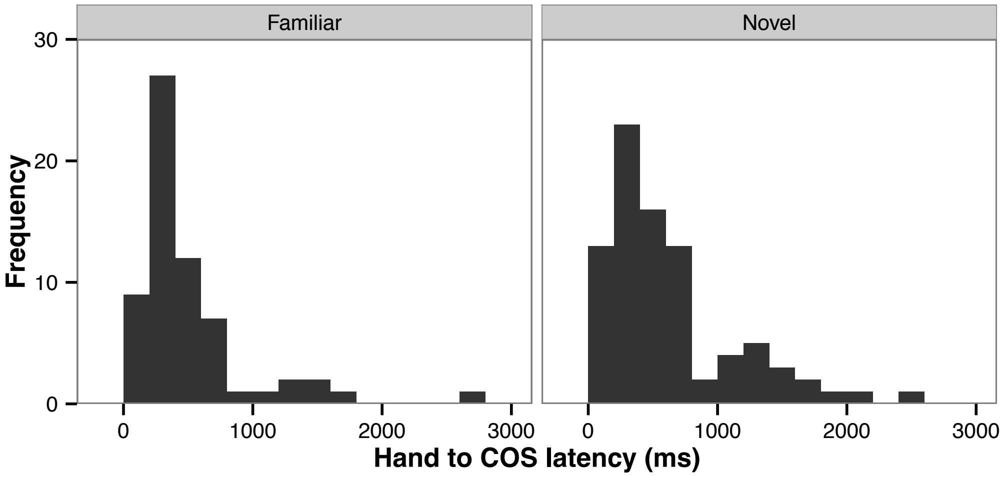


Figure 5.9 Latency between hand arriving and change of state of object

5.4 Discussion

Most environments encountered on a daily basis are familiar to us, even in novel surroundings; there are typically multiple shared statistical similarities with previously encountered environments (Ehinger, Xiao, Torralba, & Oliva, 2011). Most Western adults are familiar with domestic kitchens, their layouts and the objects likely to be contained in them, yet being familiar with the typical stereotype of a kitchen is different from having actual expertise of a specific kitchen. Having experience of kitchens in general may mean that we have some idea of the likely place that the mugs would be stored (e.g., a cupboard), whereas having knowledge about a specific kitchen would mean that we would know that mugs are stored in a particular cupboard. The prior knowledge gained from experience with an environment is likely to affect behaviour in that environment. Determining the extent to which this expertise with an environment and its objects influences our visual behaviour was the main aim of the present study.

The results indicated that depending on whether the environment is familiar or novel our visual behaviour during the completion of a well-learned task is impacted in several ways. In the familiar kitchen, participants were faster at the task and had significantly shorter eye-hand latencies. However, it is not only that performance becomes faster, the proportion of time spent performing certain types of visual behaviours change depending on whether the environment is familiar or novel. In familiar kitchens proportionately more time is spent making look-ahead fixations, less visual explorations and proportionately less time is spent ‘touching’ objects in their own kitchen than in the novel one. The present study indicates that the environment the task is conducted in affects visual behaviour even during a well-learned task.

Performance during an active task can be aided by experience of the environment the task is executed in. The results of the present study demonstrated participants were significantly faster at the same task in the familiar kitchen than the novel one. It may be that being familiar with the environment allowed participants to rely more on memory for the general global layout of the environment and the local details of the room, such as the locations of objects. Task completion may then have been facilitated by reducing search time and ensuring efficient locomotion and reducing errors such as re-inspecting unsuccessful locations for target objects. In the novel kitchen, participants had no opportunity to build a representation of the kitchen prior to the task, thus guiding fixations, particularly to objects not in plain view, could not be facilitated by remembering previous locations. In order to interact with objects in our environment that are not currently the object of fixation, we have to initially rely on our representation in order to guide the fixation into the proximity of the target object. According to Tatler & Land (2011), the egocentric spatial representation of an environment is based on the direction of the locations of objects in our environment relative to our body in space. This temporary egocentric model interacts with both visual input and the allocentric model of an environment, to allow us to depend on varying degrees upon visual information and memory (Tatler & Land, 2011). Empirically, we know that memory does indeed aid saccade planning even when we would be best served using the visual information present. In a block copying task, Aivar, Hayhoe, Chizk, & Mruczek (2005) demonstrated that when participants looked away from the resource area, saccades to blocks which had been in peripheral view but subsequently had their position changed during the look away were re-fixated from memory rather than the post change location despite being available in peripheral view at the launch of saccade. Similarly Brouwer & Knill (2007) presented two targets on screen which had to be placed in a trash bin, in some cases the position of the second target was changed when the first target was being moved. Their results demonstrated that the weighting of the extent to which participants relied on memory or visual information depended on the visibility of the second target, when the second target was lower contrast it was more common for participants to rely more on memory rather than vision (i.e. they were more likely to fixate the old location), thus indicating that the weighting system can be flexible to respond to the demands of the task and the available visual quality of the targets.

From the evidence presented so far, it would seem that in familiar settings one could either utilise memory to guide fixations or use the visually present information, or a combination of the two. If memory is utilised more in familiar environments, visual search behaviour should differ depending on familiarity. The results from the present study indicate that proportionately less time was spent visually exploring the familiar environment than the novel. Typically most of the literature examining whether people use memory or vision to guide fixations during search of familiar scenes have been conducted using photographs displayed on screens (Brockmole, Castelhano, & Henderson, 2006.; Brockmole & Henderson, 2006; Draschkow, Wolfe, & Võ, 2014; Oliva, Wolfe, & Arsenio, 2004; Võ & Wolfe, 2012). However, for information presented on screens, the cost of an eye movement is low therefore search may rely more on vision rather than memory, whereas the present study used real kitchens which had information and potential target locations spread around the room. Typically many of the target objects would be available either in the extremities of peripheral vision or even not at all without having to make a head or even trunk movement in order to fixate the target, which of course would be costly in terms of effort. Referring back to the relative weighting of reliance on memory versus visual information (Brouwer & Knill, 2007) depending on the quality of visual information available peripherally, it may be that in order to avoid having to make effortful head and trunk movements to bring a target object into clear view, search in these circumstances relies more heavily on memory in order to minimise the physical effort of searching a natural environment.

It is often the case in natural environments such as kitchens, that target objects are not just out of clear view but actually occluded either behind another object or even contained behind a cupboard door. In this way, it may be useful to consider visual search in an active task as foraging, the crucial difference between a search and a foraging task is that foraging often requires action in order for the target to become visually available, for example locomotion and manipulating objects to look inside or behind. The differences in task demands have implications regarding the cost and benefit of strategy for locating objects. According to Gilchrist, North, and Hood (2001) during a purely visual search task, relying on memory comes at the high cost of forgetting, whereas, in a foraging task, the cost of effort required to locomote and perform physical actions would be high thus reliance on memory may increase. Gilchrist et al. (2001) argue that memory plays a much greater role in search with foraging compared with visual search tasks, and that one of the motivating factors is the motivation to avoid revisiting previously checked locations, which would be unfruitful and costly in terms of time and effort. Since foraging often occurs in the absence of visual target location cues, remembering locations of previously visited locations can minimise the cost of having to make effortful full-body re-locations to conduct physical searches on locations already inspected. The notion of avoiding re-visiting locations was further examined by Smith, Hood, and Gilchrist (2008) who compared visual search and foraging behaviour within the same environmental context and found that revisit errors in the foraging condition were rare, again the authors take this to suggest that memory plays a greater role in foraging in order to avoid the increased effort required to actively search space and make revisiting errors. In fact Smith et al. (2008) go on to argue that visual search and foraging are not equivalent, particularly in terms of the scale of the stimulus, with visual search tasks typically being constrained to a screen, whereas foraging tasks typically require the participant to physically search through space. Thus the movement required to locate objects would impose not only a different temporal nature but also make revisits costly in terms of effort compared to a relatively low effort eye movement. Since the objects in the present study were not always visible without first completing a physical search, the reduction in search type behaviour (exploration) for familiar environments supports the literature on large scale search (Gilchrist et al., 2001) and foraging (Smith et al., 2008). Based on that, our findings may be further evidence that in a familiar environment, participants rely on memory, at least for the locations of objects, thus facilitating a faster performance time and reduction in time spent searching.

In the familiar environment, participants spent a greater proportion of time making Look-Ahead Fixations (LAFs) than they did in the novel environment. According to Pelz and Canosa (2001) the purpose of look-ahead fixations is to help the stream of visual input seem continuous, aid in the processing of dynamic environment where the temporal nature means that object locations may not be fixed. Mennie, Hayhoe, and Sullivan (2007) investigated the role of look-ahead fixations during a model building task and found that 20% of reaches and grasps were preceded with a look-ahead fixation approximately 3 seconds before the reach. According to Mennie et al. (2007) look-ahead fixations are purposeful and play a role in planning by facilitating the programing of the next saccade to the target object. The present finding that less time is spent making LAFs in a novel kitchen may suggest that planning several steps in advance is not possible in novel surroundings, perhaps requiring some representation that has been built up over past experiences of the space and objects in it. It may be that in a novel kitchen people are spend less time making LAFs since they would not always have the necessary information needed to locate the next target object in order to make a look-ahead fixation and after each manipulation a new visual search would have to be implemented, this would explain the higher proportion of time spent visually exploring and the lower proportion of time dedicated to making LAFs in novel environments.

Alongside the visual experience gained from familiarity with an environment, the motor actions we complete to navigate around a familiar environment can also be committed to memory. Hartley, Maguire, Spiers, & Burgess (2003) find that familiarity with a route facilitates navigation performance and speed and argue that well rehearsed motor sequences require less perceptual processing and conscious control and that we essentially formulate a representation of sequences of body movements for familiar environments, in their case well learned routes. Having a representation of the sequence of motor actions needed to make tea in the familiar kitchen may have also contributed to the reduction in task completion time. In fact, our results reveal that less time proportionately was spent during the task touching objects, this measure includes both manipulations essential to the completion of the overall task, for example pouring the water from the kettle, and those which could be considered superfluous to the actual task, for example repositioning the mug closer to the kettle several actions after its initial retrieval. Having a rehearsed sequence of movements for the familiar environment may be the root of this reduction in *touch* time. It should be noted here that the sequence of sub-goals that make up the main goal of making tea was not different in familiar or novel kitchens, this coupled with the fact that there are an irreducible set of sub-goals that must be performed in order to make tea, indicate that the reduction in touch time for familiar kitchen is more likely to be a result of utilising the representation of the movement sequence rather than a result of suboptimal sequential ordering of sub-goals being performed in the novel kitchen.

The results from present study indicate that in familiar environments, vision and action are guided to differing degrees by the representation we have of an environment, the current visual information available and by the repertoire of movement sequences we have learned and built representations for. Whether an environment and its objects are familiar or not changes the latency with which the eye guides the hand at the start of an action. Typically the eye leads the hand fairly consistently by about half a second (Land & Tatler, 2009; Land, Mennie, & Rusted, 1999), however the results of the present study reveal that in a familiar environment eye-hand latencies are significantly shorter than in a novel environment. We do not know the exact utility of eye-hand latencies other than being an example of the eye leading and guiding action, however the shorter eye-hand latencies in the familiar kitchen suggests that whatever the utility, less of it is occurring when the participant is familiar with the surroundings and associated objects. Two things could be contributing to shorter eye-hand latencies for familiar environments; firstly processes preceding the actual action may impact the latency of the eye leading the hand, for example visual behaviours and planning of actions. In a familiar environment, we find there are many more look-ahead fixations performed. Mennie at al. (2007) considered the utility of LAFs to be to aid planning the next saccade to the object and argued that objects that had previously been fixated in a look-ahead were associated with a subsequently shorter eye-hand latency, however here we find the opposite. Although the direction of effects demonstrated in the present study differ from those found by Mennie et al. their suggestion that the utility of a look-ahead fixation is planning may still be relevant for the current results. We propose that the planning reflects not only planning the fixation end point of the next object to be used, but also reflects the planning of the sequence of actions to be completed. To this end LAFs function as a checking mechanism to ensure that the action script is on track and the target objects are in location. Using these LAFs to check subsequent actions means that the requirement to have the eyes arrive well ahead of the hand is lessened and in a familiar environment since more LAFs are made, the result is a correlation with shorter eye-hand latencies. We found that across days the proportion of look-ahead fixations increase and the eye-hand latencies decrease, which is the opposite direction found by Mennie et al. (2007). We can speculate as to why this might have occurred, one possibility is that due to the nature of the two tasks, different degrees of visual guidance towards the end of manipulations may have been required and therefore the point at which the eyes could leave the current manipulation differed, which in turn effects the subsequent eye-hand latency. Although both the model building task and tea making task afforded opportunities during manipulations to make look-ahead fixations, at the end of manipulations both tasks had different constraints as to the point at which the eyes could leave. For the model building task used by Mennie et al. the end of each manipulation was screwing a nut and bolt and putting the completed model down, the authors point out that this manipulation did not require visual monitoring. Therefore, the eyes were free to look elsewhere, including the next object to be used, which would mean that the eyes would have arrived while the hands were still engaged in the previous manipulation thus producing longer eye-hand latencies. In the present study however there are two factors which may have impacted this, first several of the manipulations would benefit from continuous visual guidance (for example when pouring water from the kettle and then setting down the boiled kettle), thus the eyes were not free to fixate on the next object to be used much earlier than the hands were free to reach the next target. Second, unlike the Mennie task where both hands were required to screw together the pieces, in the present study participants could have completed an action with one hand which required visual guidance whilst simultaneously beginning the next sequential action with the other hand, thus the hand was free to leave (or arrive at the next object) but the eyes were not, which would generate shorter eye-hand latencies.

Although we do not know what is being processed during eye-hand latencies, our results demonstrate that they are affected by whether or not the participant is familiar with the environment or not. It may be that one of the reasons that eye-hand latencies are shorter when the task is conducted in a familiar environment is that during the latency some level of processing is occurring and the amount of processing required depends on how much we already know about the object. In a familiar environment, there is perhaps less impetus to have lengthy eye-hand latencies since there already exists representations of both the objects in their familiar locations and the movement sequences associated with them (Hartley et al., 2003) therefore the processing regarding the objects shape and the motor action required to interact with the object is lessened. Since no changes were made to the objects during the task the representations of objects would not have required updating with present visual information, therefore the level and perhaps length of time needed to process the object and plan the action on-line during the eye-hand latency period was reduced (Cole, 2008).

Laboratory experiments typically face time constraints that restrict the level of familiarity acquired by participants. In the real world many of the environments we encounter on a day-to-day basis are incredibly familiar, for most of us our homes, our place of work and even the route to and from our normal destinations will be more familiar than that which can be replicated in a laboratory. To exploit this level of familiarity and explore the impact it has on visual behaviour we recorded eye movements of participants performing an everyday, familiar domestic task in a familiar environment (their own kitchen) and a novel environment (another participant’s kitchen). We found that in their own kitchen participants are faster at completing the task, spend less time visually exploring the environment and make more look-ahead fixations, make less touches to objects and display shorter eye – hand latencies. Our findings suggest that not only is the overall tempo of the task overall shorter in environments that are familiar, for example in terms of task completion time, but also the microstructure of the actions in the task, for example the latency with which the eye leads the hand are shorter when a person is in an environment which is familiar to them. Furthermore the way vision is allocated during a task changes depending on the level of familiarity with the environment and objects. From the present study we can conclude that visual behaviour even during a familiar task is different depending on whether an environment is familiar or novel. What remains to be investigated is how our visual behaviour changes as we acquire familiarity with an environment and its objects.

Chapter 6 Familiarity acquisition

6.1 Introduction

The previous chapter compared visual behaviour during an automated task in both a truly novel and familiar environment. The results clearly demonstrate that being familiar with an environment changes our behaviour (both visual and otherwise): we are faster at making tea in a familiar environment, spent less time visually exploring, more look-ahead fixations were made and the latency with which the eye leads the hand during an action was shorter. Participants had lived in their homes for one year or more at the time of testing and had therefore had a considerable amount of time to acquire familiarity naturally. The process of familiarity acquisition has been examined in the literature in terms of learning an active novel task (Sailer et al., 2005) and even across an extended period of time (Foerster, Carbone, Koesling, & Schneider, 2011b) however, the way in which familiarity of environments is acquired in the context of natural task completion has yet to be examined.

When becoming familiar with an environment, several elements of our visual behaviour and task execution may change across time. It may be that as familiarity is acquired it is possible to rely more on past experience with the environment and free up some cognitive resources rather than having to process everything from scratch. In which case, we may be benefit in terms of planning or decision-making, at completing visual searches and/or at recognising/verifying the desired target objects. This can be seen in studies that have examined repeated visual search (Hout & Goldinger, 2010; Körner & Gilchrist, 2007, 2008; Solman & Smilek, 2010) where it appears that when there is a cost in terms of effort for conducting visual search anew on each presentation, (Howard, Pharaon, Körner, Smith, & Gilchrist, 2011), memory seems to come into play for aiding visual guidance, this may not be the case for simple search arrays where making an eye movement to search anew is relatively economical effort wise (Võ & Wolfe, 2011, Oliva et al, 2004). Alongside familiarity facilitating the cognitive elements of the task, the actual manipulations made with objects may benefit from rehearsal (Foerster et al., 2011a; Sailer et al., 2005).

The studies mentioned above have tended to focus on the spatial nature of gaze allocation and the effect of familiarity on visual guidance, however the temporal nature of gaze allocation has also been examined in terms of the differences in eye movements exhibited by experts compared to novices at various tasks. Prior knowledge can either be used to direct the eyes ahead of action, for example in a screen based task, in the absence of cues, viewers anticipated moving targets based on learning (Kowler et al., 1984; Kowler, 1989). Or the prior knowledge can be used to rapidly evaluate incoming information and make an anticipatory fixation at a key time ahead of action based on that information, as demonstrated in many examples from sport such as cricket, table tennis and squash (Hayhoe et al., 2012; Land & Furneaux, 1997; Land & McLeod, 2000). Typically, these studies have focussed on comparing the visual behaviour of individuals who are established experts with novices, however a few studies have since have demonstrated how eye movements change during the process of learning a new task both in a lab based study (Sailer et al., 2005) and in the real world example of learning to drive (Land and Hughes in Land & Tatler, 1999) with learners demonstrating that with a little experience they begin to be able to make anticipatory eye movements and in effect are planning ahead of the action.

Relying on prior knowledge in terms of being able to make predictions about how the world typically behaves necessitates that we have some notion of routines or events that typically co-occur or have a cause and effect relationship. Having sets of associations or world heuristics that can allow us to set expectations may also have an impact on our visual behaviour. Gaze allocation decisions can also be considered in terms of executive control, for example the *Attention to Action* model proposed by Norman and Shallice (1986) which asserts that ‘schemas’ (behavioural routines) are automatically carried out based on associations made from environmental cues. These repeated associations between environmental cues and the related response behaviour are, according to Norman and Shallice, pervasive and bias action toward certain behaviours in familiar environments, however this schema activated behaviour can be overruled by the ‘supervisory attentional system’ in situations where the schema fails to meet the behavioural goals. Thus being able to direct gaze to informative areas of the scene, particularly during dynamic events may be benefited by employing more general prior knowledge about the way the world works. However, it may be that in certain circumstances relying on prior knowledge is actually a hindrance. Hangovers of certain behaviours may in fact impair one’s ability to respond to a change in circumstance. A classic study conducted by Luchins (1942) where observers demonstrated a predisposition to solve a given problem in a specific manner even though superior or more appropriate methods of solving the problem existed revealed the negative effect of previous experience when solving new problems, this reliance on inefficient strategy was referred to by Luchins as the '*einstellung effect',* and was considered as the development of a mechanized state of mind. This suggests that schemas are not constantly evaluated and in fact behaviour may not be optimal. In terms of the extent to which schemas are built up in natural sequential tasks and whether there is an effect on planning our visual guidance has yet to be examined.

Planning where to look next when performing a visual search in order to find a target object can be influenced by the level of previous experience with a scene or object. Each time a new object needs to be located, there is the option to either conduct a visual search, use memory of the objects location to drive the gaze allocation or integrate both sources of information. In cases where we have no specific knowledge about where an object is, we can guide the search using expectations based on our general prior knowledge of the world. Visual search may be facilitated or constrained by factors such as scene context: Neider & Zelinsky, (2006) demonstrated that observers will search a scene differently depending on the target of their search and their expectation of where this target might appear in the scene. Similarly Kanan et al. (2009) argues that viewers make extensive use of knowledge about where and how objects tend to appear in a scene and use this information to guide search. Typically in natural searches, we not only have experience of where certain targets are likely to be depending on the context of the scene, it is often the case that tasks are conducted in environments where we have specific experience of the layout of objects. In scene viewing, there is strong evidence that searching for an object the second time is marked by significant reductions in search times and fixations to irrelevant locations, Hollingworth, (2012) found that visual search is speeded after familiarization with a new scene whilst Võ and Wolfe (2012) demonstrated that visual search is speeded up when looking repeatedly looking for an item even despite intervening irrelevant searches. Typically however the literature on this issue has tended to use static displays only, where normally one would expect the layout of the scene and the objects contained in the scene would remain constant, when in contrast in real life the dynamic characteristics of the environments we perform tasks in may complicate the issue somewhat, for example some aspects of real environments are constant but others are subject to change, objects often move whereas room fixtures rarely do. Therefore although it is possible to utilise general semantic guidance of appropriate places to look for target objects, there must be some flexibility in order to cope with the more dynamic element of natural environments.

From the literature presented above, we know that we bring our prior knowledge about the world to tasks and in certain circumstances use it to guide fixations, for example if we were searching for a kettle in a kitchen, we would utilise what we know about where things are usually kept in kitchens and perhaps restrict our search to a worktop near an electricity outlet. This type of prior knowledge is fairly general and the type that all normal adults build up through the course of normal daily life. Similarly from scene viewing experiments we know that having gained experience of a scene from repeated viewing and searching, subsequent searches are facilitated (Vo & Wolfe, 2011). Another more specialised type of knowledge that can influence visual behaviour is expertise, studies comparing novices and experts have found that the two groups display different visual behaviour for the same tasks, with experts seemingly able to utilise their experience to facilitate guiding the eyes ahead of the action and that eye movements appear to change as a task is learned. We also know that during the process of learning a task, visual behaviour changes, both in terms of becoming more proactive and anticipating action (Sailer et al., 2005) and with regards to the microstructure of the ORA (Foerster et al., 2011).

What we do not know from the literature is whether acquiring familiarity with an environment and the contained objects would also produce changes in visual behaviour in a similar way as occurs for task familiarity. This is an important question to address since much of what we actually do in the real world is repetitive, unless we are learning a new job or hobby we typically perform tasks that are already familiar to us, whereas in a typical day we may face several new environments, for example if we visit a new friend, or restaurant or take a new route to work. Considering the significant impact acquisition of familiarity with a task has on our visual behaviour it is worthwhile investigating the effect of having to perform am familiar task in a novel environment and whether visual behaviour changes as familiarity with the environment is acquired.

6.2 Method

6.2.1 Participants

Ten undergraduate students (2 male) from the University of Dundee participated in the study on a voluntary basis. All participants had normal or corrected-to-normal vision.

6.2.2 Materials

A kitchen in the University of Dundee, School of Psychology building was utilised for this study. The room fixtures included a worktop and kitchen sink, a shelf and electricity points, a fridge and under worktop cupboards. Task relevant objects, i.e. those for making tea, were laid out in appropriate locations throughout the kitchen. Several of the objects required for making tea were located in cupboards. The kitchen also contained distractor objects all of which were items typically found in a kitchen, for example glasses, dishes sponge, cutlery, plates and toaster. For all ten days of the task there were also present the objects required for Chapter 7, two sets of objects (items to make a peanut butter and jam sandwich and a mug of fresh coffee) were located in similar positions to the tea making objects and as far as participants knew for the purposes of the present study were simply more distractor objects. Perishable objects for the task were laid out in the kitchen and replaced as and when necessary.

6.2.3 Design

We used a within subjects design with all participants tested across the 10 day period. Objects were always in the same place before the start of each testing session and participants were given the same instructions every day, they were not given any instruction as to the order of sub goals to be completed.

6.2.4 Procedure

On day 1, after providing written consent to participate in the study, participants were informed that they were required to make a cup of tea in a real kitchen, using a floral teacup, a teapot and milk. It was explained that just like in any other kitchen the objects they would need would be dispersed throughout the kitchen and that they were permitted to look anywhere. It was stressed that as much as possible we would like them to be entirely natural and make the tea just as they would should they be making it for themselves, and in fact participants were encouraged to drink the tea after the end of the experiment in order to encourage them to make the tea properly.

Participants were then calibrated using the calibration system detailed in the general methods chapter. This procedure took place in the hallway outside of the kitchen, in part to minimise the chances of the participant viewing any of the kitchen but also to maximise the wearing time of the eye tracker before the trial commenced. After calibration the instructions were briefly recapped and the participant's comfort in wearing the eye tracker were re-checked. Participants were reminded that there was no time pressure and that they were to complete the task in the way they would normally do so (i.e. there was no set order for the sub components of the overall task); participants were given no instruction to tidy up at the end of the task. The procedure on all 10 days was the same, minus the detailed instructions at the start and the consent signing.

6.2.5 Analysis

We were interested in several different measures of eye movements and behaviour, which can be broken down to the following categories.

6.2.6 Eye Hand latencies start of manipulations

Eye hand latencies have frequently been measured and reported as a microstructural element of the ORA (Foerster et al., 2012; Foerster et al., 2011; Land et al., 1999; Land & Hayhoe, 2001) several studies have identified that a feature of vision during an active task is that vision is proactive and the eye tends to lead the hand by roughly a second. These latencies have been demonstrated to change as familiarity with a task was acquired. For the present study, the latency between the eye and hand arriving were analysed at the beginning of each manipulation across the 10 days.

6.2.7 Eye and Hand Latencies at the start of manipulations to Change of State

The time between the eyes first arriving on the object and the actual point in time where the state of the object changes was measured and analysed across both trials and days, similarly the point where the hand touches the object and the change of state occurs was measured and analysed. The correlation between the eye hand latencies and the hand to change of state latencies were also analysed across days. The delay between the eye arriving and the beginning of a COS and the hand arriving and the change of state may vary as a result of familiarity, which may indicate whether familiarity effects the need for processing information about the functionality of the object and planning of the manipulation during this phase of an ORA.

6.2.8 Change of State (COS)

The total time for a change to state to be completed was analysed for instances where the manipulation lasted longer than a frame (on many occasions, for example the difference between a cup 'being down on the worktop' and 'up' would last for only 1 frame).

The time between the end of the COS and the eyes leaving and also the end of COS and hand leaving were also measured and analysed. Correlations between the total COS times from start to end and the eye hand end latencies were also analysed. Measuring the time taken to complete the change of state gives us an indication of whether any reductions in task completion time could be attributable to faster completion of the manipulations.

6.2.9 Put downs

The way objects are treated at end of manipulations are interesting because as an object is put down on a surface it is a clear interaction between the object and the environment. To accurately set an object down on a surface without breaking or spilling or colliding with another object it would seem intuitive that one would always use visual guidance to ensure an intact putdown, however we know that this does not always happen, some putdowns of objects are visually guided and some not. The results from Chapter 4 suggest that the level of visual guidance directed towards an object during the process of putting it down on the worktop varies depending on the properties of an object and the level of clutter of the set down surface, in order to examine whether there was an effect of acquiring familiarity with the environment and the objects in it. The way objects were visually treated after a manipulation had been performed on them was measured by recording whether each put down of the object was visually guided or unguided. Behaviourally we also measured and analysed whether unguided putdowns happened more frequently for some objects than others.

6.3 Results

6.3.1 Overall task completion time

By the end of the ten days all participants were faster at completing the task than on day one as can be seen from Figure 6.1. It is worth noting the discrepancy in the pattern for participant three on day three. In this instance the participant filled the kettle up and boiled it from cold which differed from the usual protocol of not filling and re-boiling a pre-boiled kettle, hence the increased task completion time representing waiting time for the kettle to boil. This participant was subsequently removed from analysis due to the deviation in task procedure.

The overall time taken to complete the task decreased significantly across days. LMM analysis revealed that the task was completed approximately 10 seconds faster per day ($\beta = -10.185$, $SE = 1.642$, $t = -6.202$, $p = <.001$). However, a model that took the log of day ($\beta = -47.42$, $SE = 8.44$, $t = -5.62$, $p < .001$) provided a significantly stronger fit to the data, assessed by comparing the model fits using an ANOVA (χ^2 (0) = 28.17, $p < .001$).

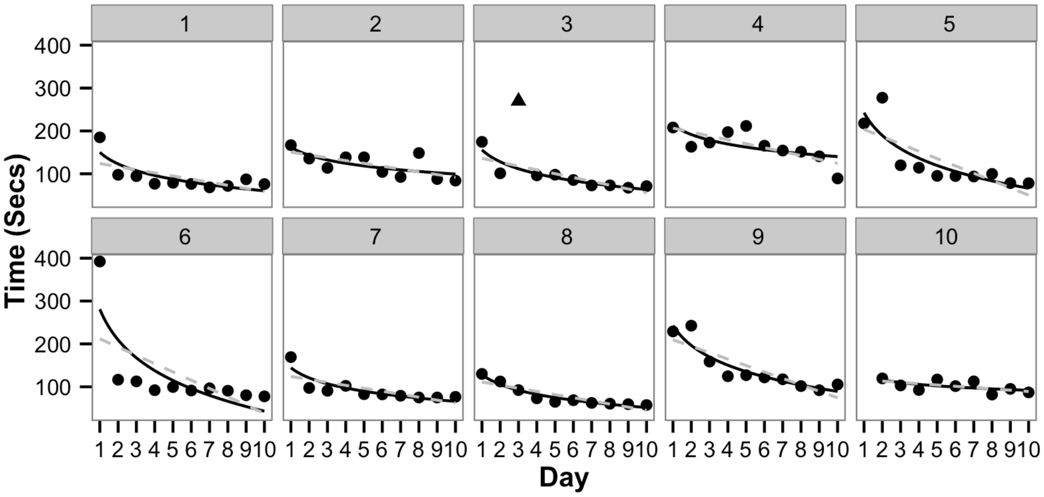


Figure 6.1 Overall task completion times per day for all ten participants. Each facet represents an individual participant. Linear model – dashed line, log(Day) model – black line. Note participant 3, day 3 was not included in the modelling, but is shown here as a filled black triangle to demonstrate their outlier.

6.3.2 Time spent on each element of behaviour

6.3.2.1 Duration of the separate elements of behaviour across days

We know that the overall time to complete the task reduces across days but it is not clear where that time is lost. Here we separate the trial into four categories or sub-elements of behaviours (eye-hand latencies, exploration, look-ahead fixations and time touching objects), in order examine whether the time spent during the task on these sub-elements changes as familiarity is acquired. To do this we ran a LMM with log(day) as the fixed effect, and participant as a random effect.

There was a significant log-linear relationship between day and the duration of the trial spent during eye-hand latencies (β = -6.32, *SE* = 1.04, *t* = -6.083, *p* <

.001), exploration (*β* = -17.29, *SE* = 3.04, *t* = -5.69, *p* < .001) and touching/manipulating objects (*β* = -16.78, *SE* = 4.89, *t* = -3.434, *p* < .001). There was no significant relationship between the sum duration of look-ahead fixations and days (*p* = .07)

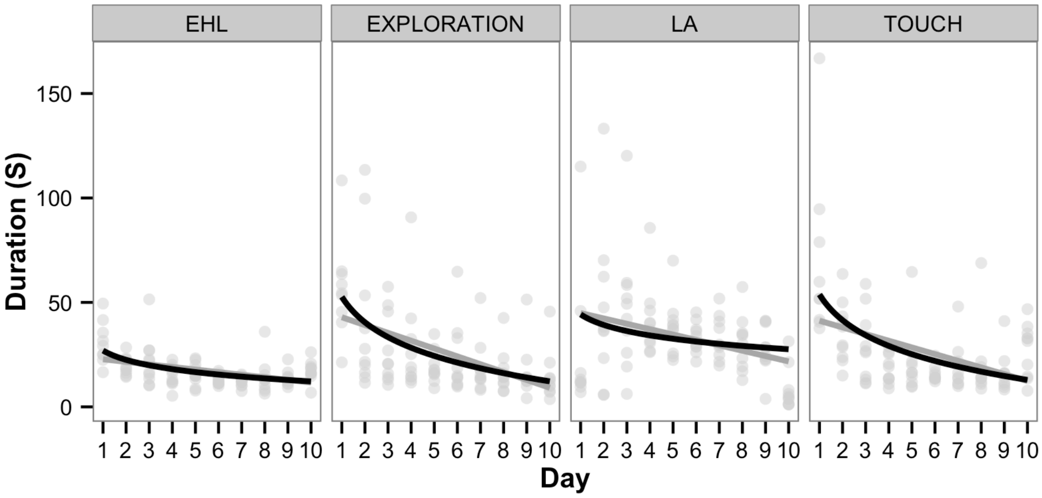


Figure 6.2 Duration within each day that was spent on each of the four sub-visuo-motor routines. Individual data is represented by dots, with log model (black line) and linear model (grey line).

6.3.2.2 Relative proportion of the separate elements of behaviour across days

One potential confound in these results is that the overall time that it took participants to make tea decreased over days (see results above). To account for this reduction, we expressed each measure as a proportion of the testing session spent completing this visuomotor routine. Thus, each participant had a sum for eye-hand latencies, look-ahead fixations, searching and touching objects of 1, which allowed us to examine whether the proportion of time (i.e. a breakdown of the task) changed across days.

There was a significant negative log-linear relationship between day and proportion of time spent exploring (*β* = -0.05, *SE* = 0.02, *t* = -2.86, *p* < .001), and a significant positive relationship between log(day) and the proportion of time looking ahead to objects (*β* = -0.05, *SE* = 0.02, *t* = 2.23, *p* = .03). However, there was no change in relative proportion of the testing session spent on of eye-hand latencies (*p* = 0.1) and relative proportion of time spent touching objects (*p* = 0.33).

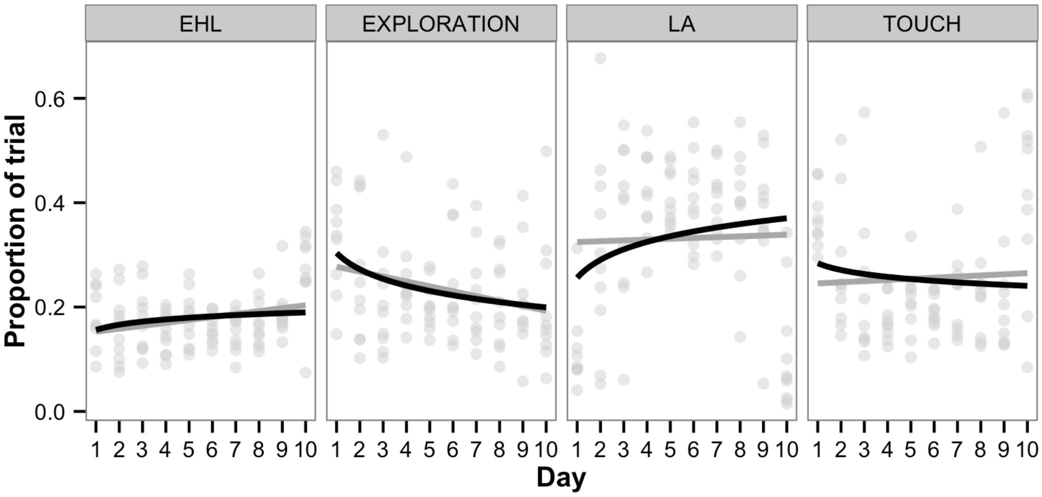


Figure 6.3 Relative proportion of the trial spent completing each of the four sub-visuomotor routines.

6.3.3 Looks to task relevant and task irrelevant objects

The proportion of time spent looking at different categories of objects in the room changes across days, the nature of the changes are plotted in is demonstrated in the figure below.

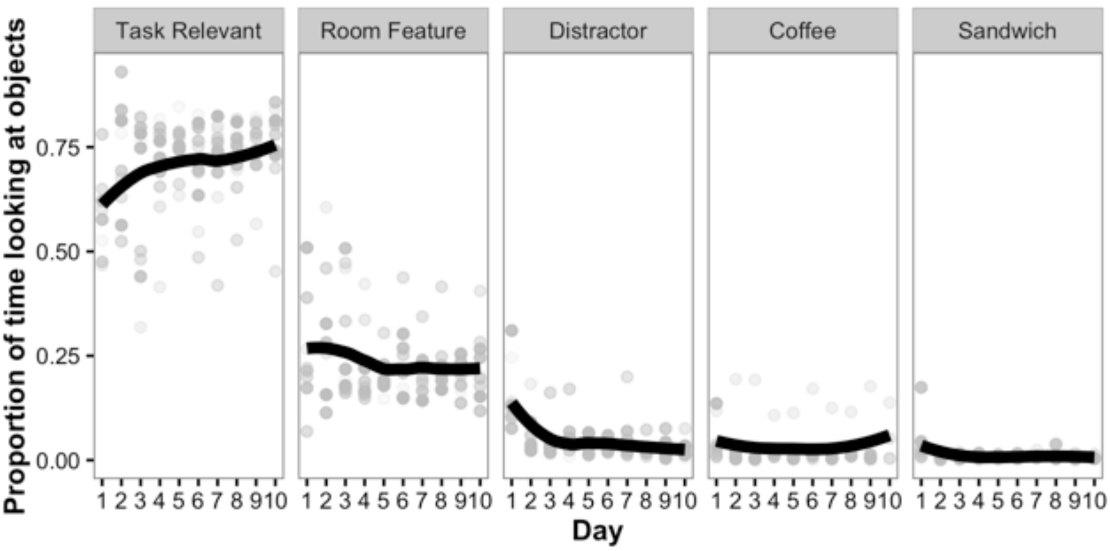


Figure 6.4 Proportion of time spent looking at different object types across all ten days for all participants.

An LMM was conducted with the proportion of looks to relevant objects as the outcome, the log of day as a fixed effect, and participant as a random variable. There was a significant log-linear increase in the proportion of looks allocated to task relevant objects across days (β = 0.14, SE = 0.02, t = 7.2, p < .001).

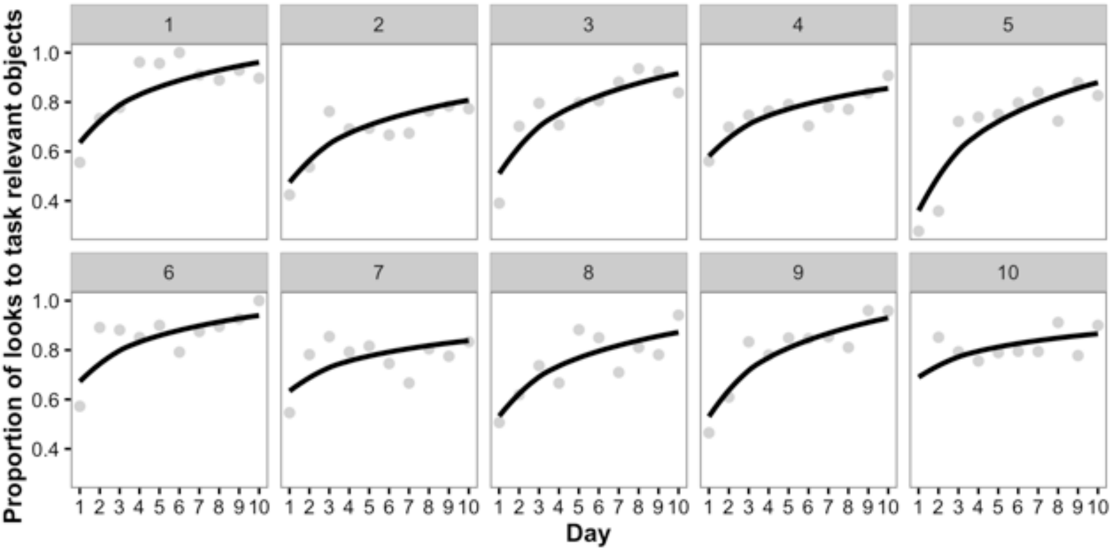


Figure 6.5 The proportion of looks at task relevant objects plotted across days. Participants are shown in different colours with a Loess best-fit curve across the whole dataset.

One possible confound in the data is that while there might be a lower proportion looks to task relevant objects, that these looks might be longer in duration. To discount this possibility the relationship between the total duration of looks and the frequency of looks to different object categories is plotted in Figure 6.6 below. If this confound existed, we would expect that as the total gaze duration increases, that task irrelevant looks should decrease. However, there is a clear positive relationship in all of the object categories, suggesting that this was not the case.

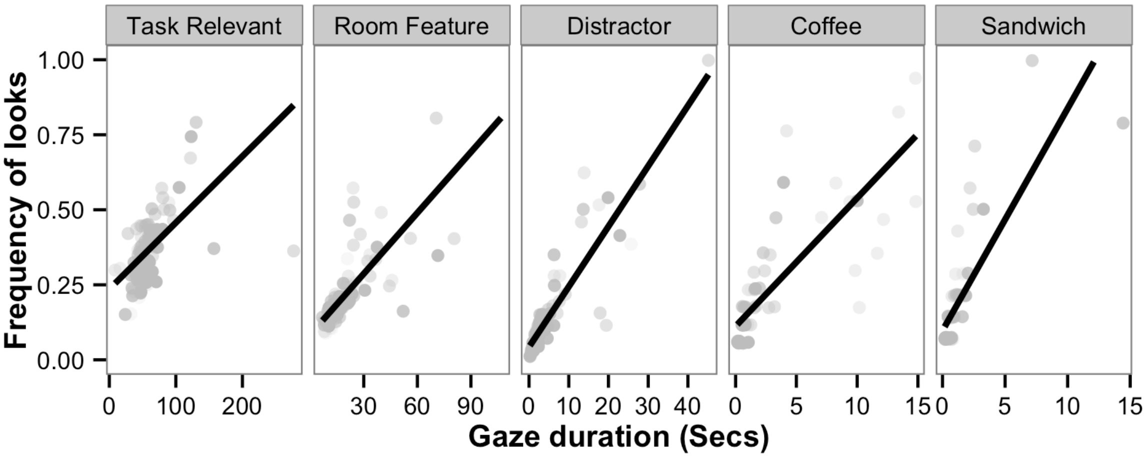


Figure 6.6 Relationship between total gaze duration and frequency of looks to objects (normalised to the maximum number of looks for easier interpretation) to demonstrate no trade-off between the two variables. Individual data are presented as shades of grey, with data jittered for clarity.

6.3.4 Eye hand latencies (beginning manipulations)

The duration of each of the eye-hand latencies that participants made when interacting with objects was analysed. Eye-hand latencies across all participants and all days are shown in Figure 6.1. Overall, the median latency between the eye arriving and the hand arriving across the entire experiment was 620.7ms. Confirming previous findings (e.g. Land & Hayhoe, 2001), in the majority of cases (95.9%) the eye leads the hand, with far fewer instances where the hand arrives before the eye (2.99%) and the remaining 1.1% of the data being instances when the eye and the hand arrived simultaneously.

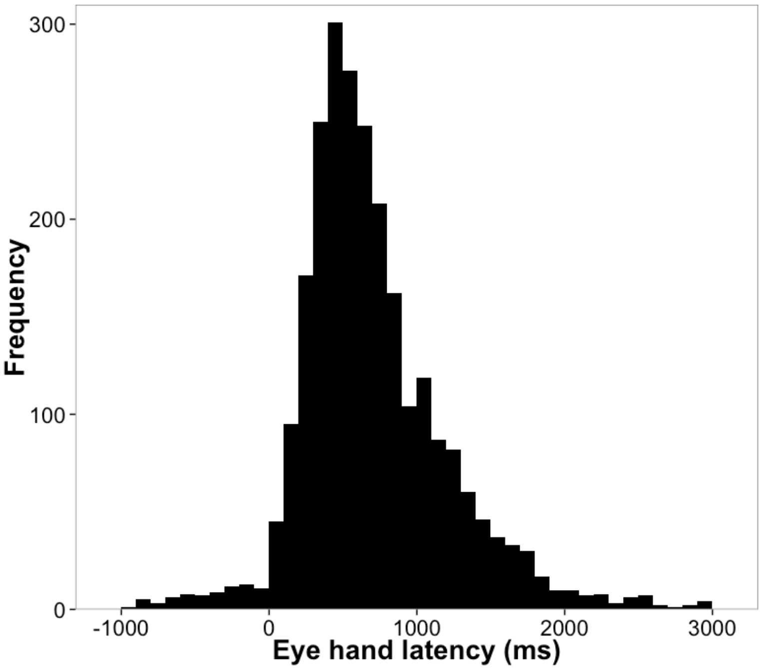


Figure 6.7 Frequency of eye hand latencies across 10 days for all 10 participants

A LMM analysis was used to examine how eye-hand latencies changed across days, with participant and the object being manipulated included as random effects. There was a significant reduction of approximately 13 ms per day (β = -12.658, *SE* = 3.757, *t* = -3.369, *p* = .005). However, the loglinear model provided a significantly stronger fit to the data (β =-49.66, *SE* = 15.66, *t* = -3.17, *p* < .001). The nature of the change across days can be seen in Figure 6.8.

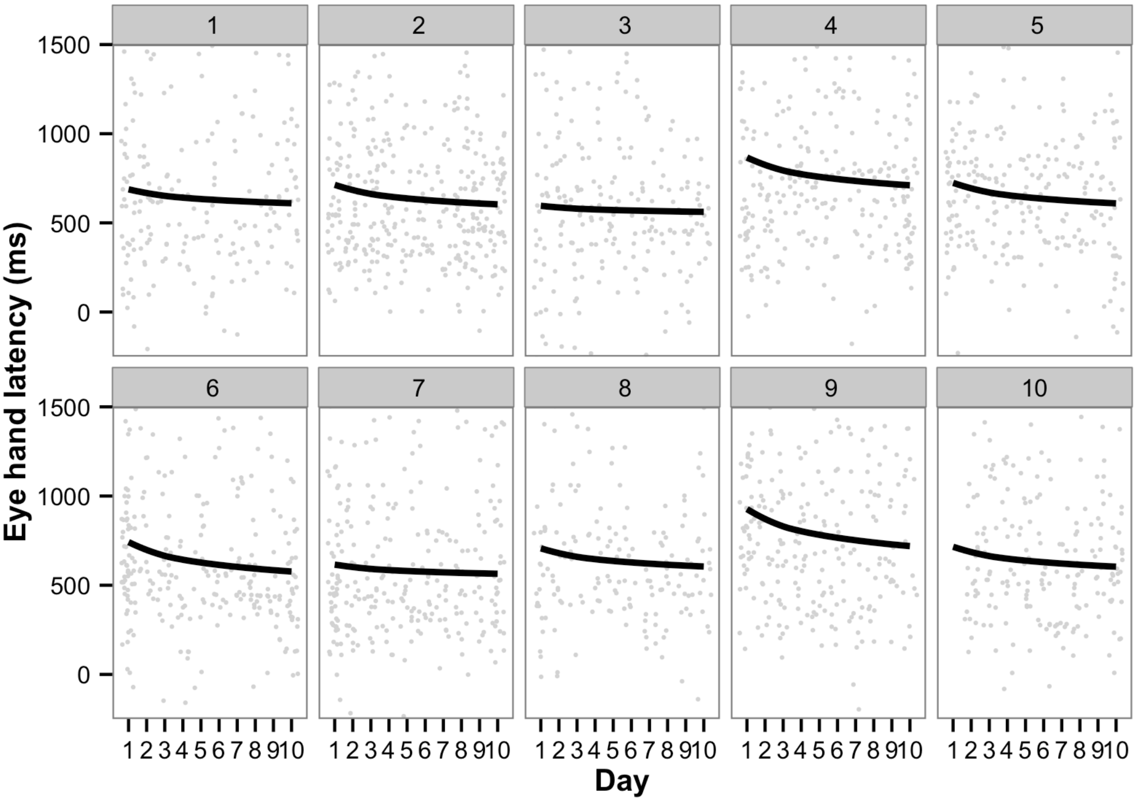


Figure 6.8 Eye hand latencies for all participants across days adjusted in the LMM with loglinear fits for each participant (represented by the black lines in the separate facets). Raw data is shown by the grey dots. For visibility, the y-axis being trimmed to -250-1500ms, although the LMM fits are taken from the whole data. For similar reasons, data-points are jittered on the x-axis to avoid points overlapping (and therefore hiding the true data).

In Chapter 3, we observed differences in eye-hand latency depending on the object being manipulated (justifying our use of LMM analysis with object being included as a random effect). We therefore examined whether the object type had any impact on eye-hand latencies in this experiment by removing the random object term from the eye-hand latency LMM model and examining whether this changed the model fit. There was a significant reduction in the amount of variance accounted for by the model when object was removed (χ^2

(3) = 385.4, *p* < .001 Figure 6.4 demonstrates that certain objects incurred longer eye-hand latencies than others, with the longest latency for kettle and the shortest being for milk lid. Figure 6.9 shows the eye-hand latencies for day 1 and day 10.

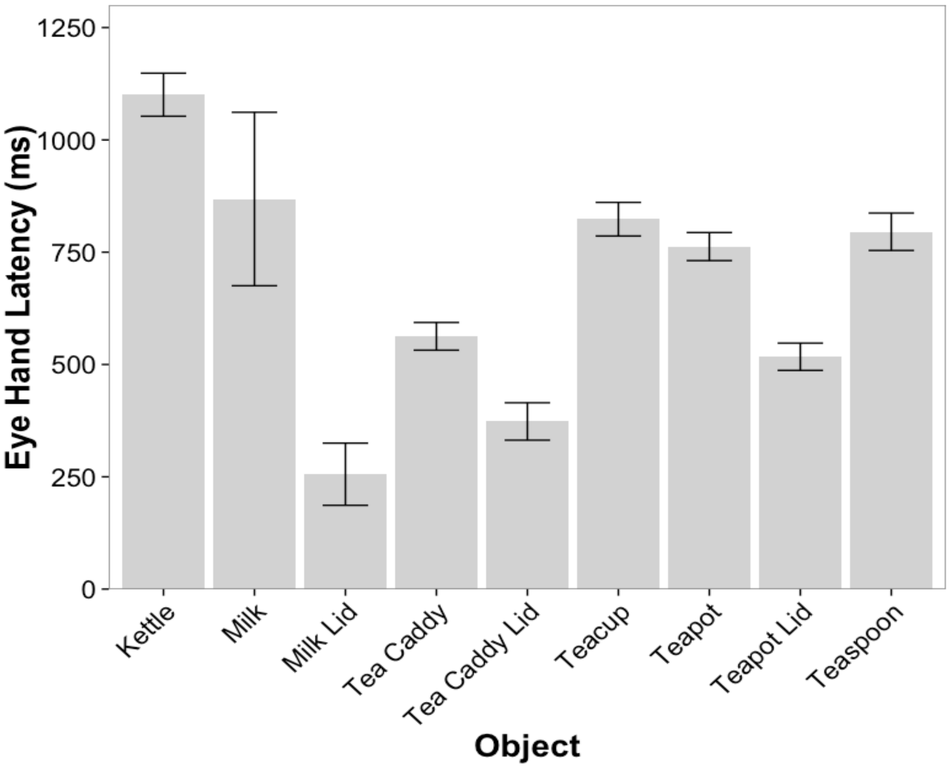


Figure 6.9 Eye hand latencies for all participants across all ten days by object type.

6.3.5 Eye to change-of-state latencies

The time between the eyes arriving at an object to be manipulated and the point in time where there is a change of state of that object was analysed using a LMM with object and participant included as random effects. The results demonstrate a significant loglinear reduction across days (β = -78.07, *SE* = 24.73, *t* = -3.16, *p* = .008). The nature of these reductions are presented in Figure 6.10.

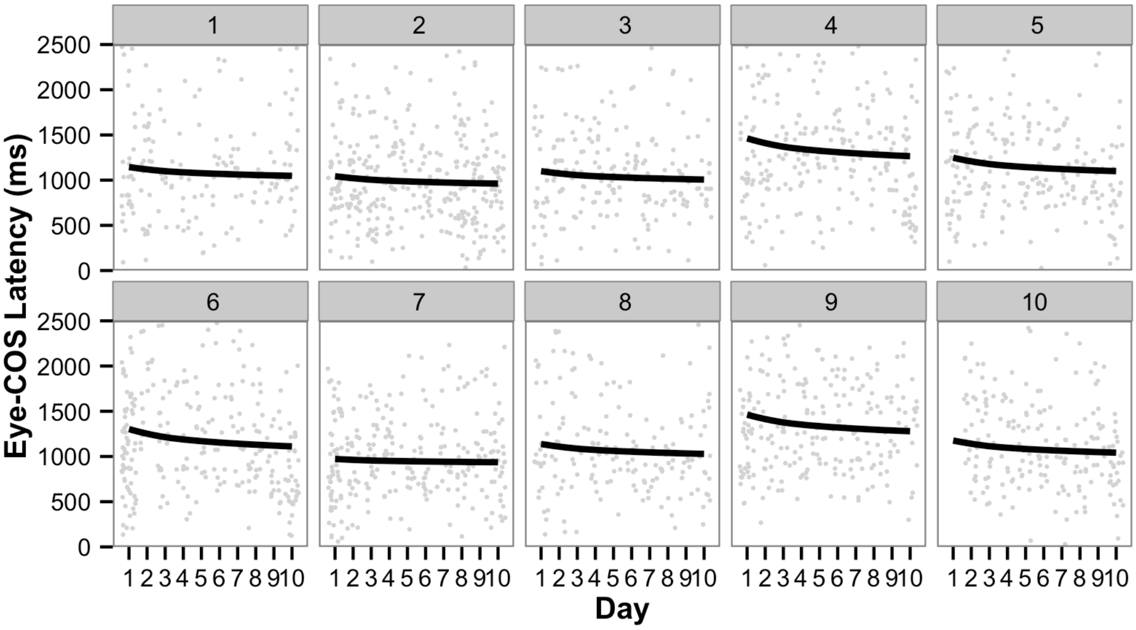


Figure 6.10 Eye to COS latencies for all participants across all ten days adjusted in the LMM with individual facets showing participant level data with LMM loglinear fits.

6.3.6 Latency of hand arrival on object to change-of-state beginning

The length of time between the hand reaching the object and the beginning of the COS was also analysed, with object and participant included as random effects. The LMM analysis revealed a significant reduction of approximately 11ms per day (β = -11.063, *SE* = 4.137, *t* = -2.674, *p* = .014). A LMM using the log of day did not converge. Figure 6.11 demonstrates this change across all testing days.

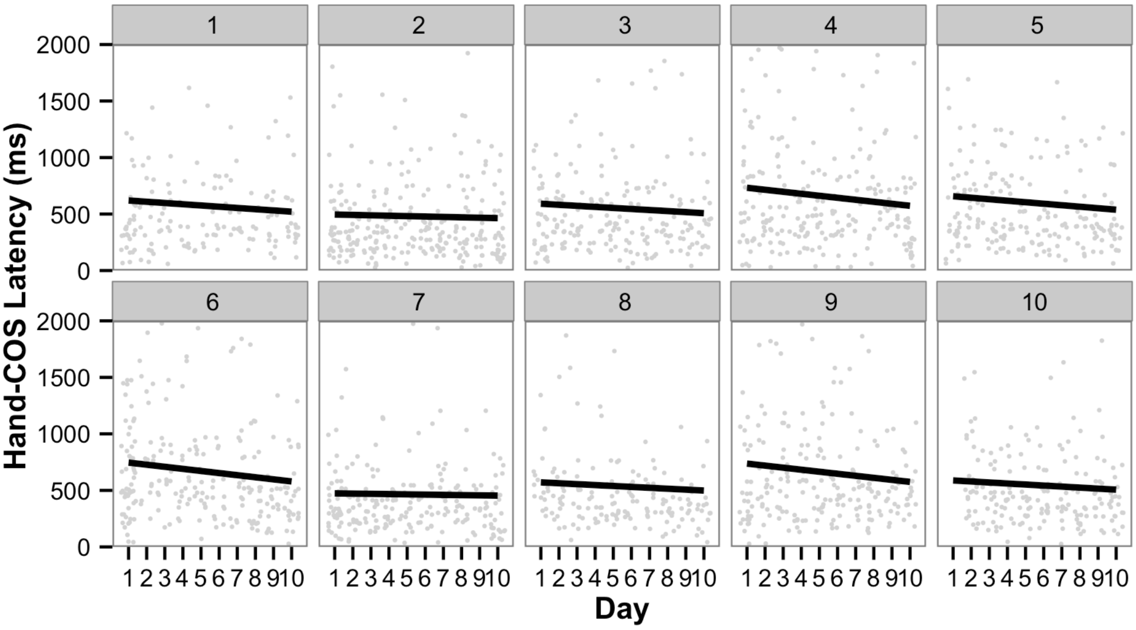


Figure 6.11 Hand change of state latencies for all participants across all ten days adjusted in the LLM. The average intercept was 632.35 with a decrease of 11.06ms per day.

6.3.7 Change-of-state completion time

The overall time taken to complete the manipulation (COS start to end point) also reduced across the ten days: LMM analysis revealed a reduction of approximately 30 ms per day (β = -30.55, *SE* = 14.44, *t* = -2.116, *p* = .049). Again, there was no improvement using the log function of day, with the model using the linear relationship providing a significantly stronger fit to the data (χ^2 (0) = 8.6, *p* < .001).

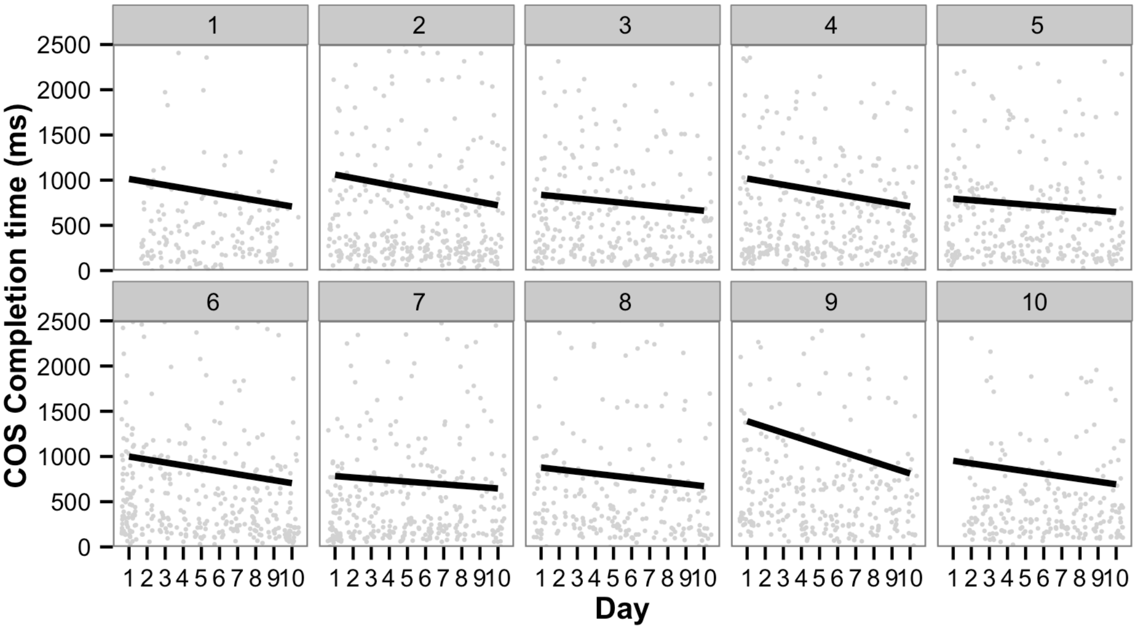


Figure 6.12 Change of state completion times for all participants across all ten days adjusted in the LLM. The average intercept was 632.35 with a decrease of 11.06ms per day. For visibility the axis is trimmed at max 2500ms, although the LMM was fitted across the whole dataset.

6.3.8 Visual guidance of putdowns

A GLMM was used with whether the object pickups and putdowns were guided (1) or unguided (0) as the binary outcome variable. The fixed effect was day, and the random effects were participant and object being picked up, or put down as random effects. The maximal model did converge for object pickups, but was not significant (*β* =-0.86, *SE* = 0.7, *z* = -1.22, *p* = 0.22). When examining object putdowns, the initial maximal model did not converge, so the coefficients between slopes and intercepts were removed. There was no significant relationship between day and the probability of objects being guided (*β* =-0.001, *SE* = 0.034, *z* = -0.03, *p* = 0.98). Thus, guidance in the picking up, and setting down of objects did not change across days.

6.4 Discussion

The acquisition of familiarity with an environment and its objects produces changes in both general behaviour and visual behaviour. The present study required participants to complete the same well-rehearsed task of making tea for ten consecutive days in the same environment. We found that across days not only did people take less time to make tea, action sequences, i.e. the order in which sub-goals were completed got more similar to both the previous days script and to other participants. For visual behaviour, the way we assign our fixations also change, both in terms of what we look at and the related timings of these fixations. Overall we can conclude that more experience with an environment and objects changes the way people visually behave during the completion of an active familiar task, and there is a progression of change from an initial exploratory visual approach to a more direct one that appears to involve more planning. This may reflect the increased ability to rely on memory to guide fixations or it may simply be practice effects of having worked in the environment for a number of days, the nature of changes and the potential implications regarding what they reveal about visual behaviour as familiarity is acquired will be the focus of this discussion.

6.4.1 Reductions in overall task completion time

Acquiring familiarity with an environment affords the benefit of becoming faster at even a well-learned task. The overall time taken to complete the task gets significantly shorter across the ten task days, reducing by approximately a third. This is in line with one recent study where the impact experience has on both motor and visual behaviour during a manual task was investigated by Foerster et al. (2011). Their results demonstrated that over the course of numerous repetitions of a bi-manual cup-stacking task, overall completion time reduced significantly and fewer fixations were made on the last training day compared to the first. One possible interpretation of their results is that the time reduction could simply be down to practice, the *power law of practice* dictates that as we become more experienced with a task performance speeds up in a log-linear fashion (Newell & Rosenbloom, 1993). In a classic study demonstrating reductions in task completion time due to task rehearsal Crossman (1959) found that the manufacturers of hand rolled cigars got significantly faster across several years (in the course of rolling 20 million cigars) until the speed levelled out as the physical limitations of the task dictated.

6.4.2 Reductions in manipulation time

The present study finds along with the reduction in overall completion time, the overall time from the start of a change of state (COS) to the end of a COS of an object reduced across the ten days, so people were faster at performing the manipulations of objects, much like was demonstrated by Crossman (1959). This is interesting since the task was a familiar task and used objects that are typical of the task, however we know from Chapter 4 that the properties of objects influence our behaviour with them so perhaps something about practising with these specific objects speeded up performance. As familiarity is acquired it could be that we use our experience with objects and the actions made during the associated manipulations to aid subsequent interactions. Although the objects used in the present study were not novel object types, participants had not previously seen them and certain objects had a quirky functional element. For example, the teapot lid had a notch, designed to hold it in place, which had to be aligned with a corresponding groove in the teapot. Therefore as experience is gained with certain objects perhaps the COS is faster to complete since we know the exact details of object properties and can rely on that to guide our interaction rather than just our representation of *all teapots.* Much like is found for visual based tasks where participant become highly skilled at recognising objects after training on the specific objects, referred to a stimuli specify, but these training effects do not transfer to other objects, even ones of the same category (Ahissar & Hochstein, 1993; Fahle, Edelman, & Poggio, 1995; Furmanski & Engel, 2000). Learning not only functional quirks of objects but also the limits set by objects mechanical constraints during manipulation may be an advantage of acquiring familiarity which improves task performance in much the same way as task practice effects.

The period of time where both the eyes and the hand arrive and the manipulation of the object begins can also be considered as a latency which may be affected by the acquisition of familiarity. Across the ten days the latency between the eyes arriving and the object changing state significantly reduced, as did the latencies of hand arriving and the following change of state. Therefore, we can say that as familiarity is acquired individuals are quicker to initiate changing the state of the object in hand after the arrival of both the eyes and hand to the target object, perhaps just as the actions themselves speeded up, so too did the time taken to initiate the change of state after the respective arrival of the eye and hand. Quite simply, it may be that practice of the physical manipulations performed for each object facilitates faster performance not just for the start of the change of state to the end (i.e. the manipulation) but also for microelements of the task such as the latency between eyes and hand arriving and the initiation of the change of state. Foerster, Carbone, Koesling, & Schneider (2011) also speculated that speeding up of the time for the eye to arrive and the initiation of the COS may occur and referred to it as the possibility of the eye – hand co-ordination becoming more dynamic with practice.

6.4.3 Accounting for the time spent on behaviour types during the task

The results from the present study indicate that behaviour changes across days, if we group behaviour into two categories visual behaviours, those that do not involve action, look-ahead fixations and exploration*,* and those directly linked with action, *touch (all interactions which involve action/object manipulation)* and *eye-hand latency,* we can consider that these changes in time spent on each behaviour type across the days likely come from acquiring familiarity. When the time spent on each category was examined, it was not the case that each of the four categories got equally shorter across the ten days. The amount of time spent performing each of the categories across the 10 days was calculated both proportionately and absolutely. In absolute terms, significantly less time was spent across the ten days performing exploration, with the eyes leading the and on commencement of a touch (eye hand latencies) and touching objects in general. The amount of time spent performing look-ahead fixations did not differ significantly across days, as illustrated in Figure 6.13.

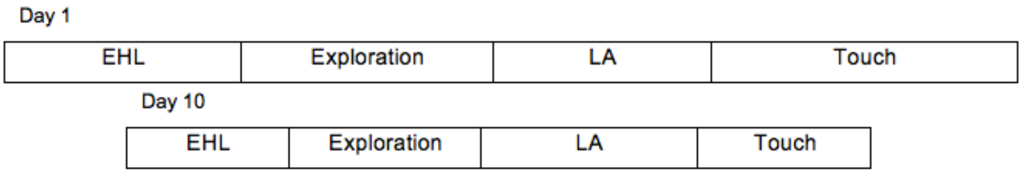


Figure 6.13. Visualisation of total time of each testing day spent on the four behavioural elements on day 1 and 10.

Proportionately, significantly less time was spent visually exploring the room and its objects across the ten days, however the proportion of time spent making look-ahead fixations significantly increased, as can be seen in Figure 6.14 below.

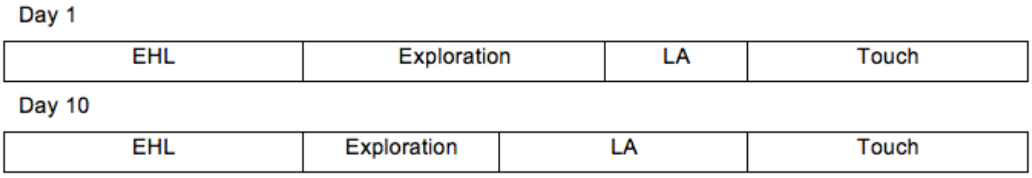


Figure 6.14. Proportion of time spent engaged in each behavioural element.

6.4.4 Proportion of time spent visually exploring the environment

As we become more familiar with an environment it may be that our representation accumulates more detail and as such, we improve at dealing with the unfamiliar environment. One of the ways that familiarity seems to aid our performance is in terms of reducing task completion time, and in particular, as can be seen from the figures above, experience may reduce time spent on certain elements of the task. Time spent on visual exploration reduces across the ten days, which may indicate a greater reliance on memory for objects locations. Similar evidence has been demonstrated in the laboratory: Vo & Wolfe (2011) presented a visually based search task and found that if a target had been previously searched for, subsequent searches were significantly faster, concluding that prior search facilitates subsequent search. We already know that task expertise changes the way we guide our fixations, for example experienced clinicians are faster at directing their gaze to diagnostically informative areas than students when viewing a virtual microscope slide (Krupinski et al., 2006), similarly, there are several examples of experience facilitating anticipation from driving (Land & Tatler, 2001; Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003a, 2003b) and in sport (Hayhoe et al., 2012; Land & McLeod, 2000; McKinney, Chajka, & Hayhoe, 2008). However our results indicate that it is not only task expertise that changes the way we allocate our gaze, expertise with the environment and its objects can also reduce search and potentially aid planning of future task elements.

6.4.5 Proportion of time spent looking at task irrelevant objects

As an environment and its objects become familiar, there are less looks to and time spent looking at task irrelevant objects indicating that the quality of representations of the environments has improved so that fixations are used deliberately to concentrate on extracting task relevant information. Just as familiarity with the environment impacts the frequency of task irrelevant looks, the physical details of the environment also influence opportunity to make task irrelevant looks, particularly if the target objects are spread out and concealed. Analysis of the present studies results revealed that across the ten days the total number of looks to task irrelevant objects decreases, whilst the duration of the gaze event on an object also decreases. Typically on day one, the split of looks to task relevant and to irrelevant objects was roughly even, whereas by day 10 only around a fifth were to task irrelevant objects. The proportion of looks to task irrelevant objects is much higher for the present study than those previously found by Land, Mennie, & Rusted, (1999). Hayhoe, Shrivastava, Mruczek, & Pelz (2003), noted in one version of the sandwich making task where the objects for making tea were located amongst other task irrelevant objects, initial fixations, before task instructions were provided, revealed that fixations between task relevant and irrelevant objects was split almost evenly (48% on irrelevant objects) but during task performance these reduced to around 16% which is clearly very similar to those found in the present study. Our results revealed a much higher proportion of looks to task irrelevant objects than in Land, Mennie & Rusted's (1999) original tea making study, despite both using the same task of making tea. In order to try to explain the discrepancy then, it is important account for differences between the two tea studies. In the original tea making study (Land et al., 1999) most of the objects required for use were within plain view, typically all of the objects (except from the milk) were available on worktops and open shelving, whereas the present study attempted to replicate a domestic kitchen setting where several of the items were located in cupboards, which of course necessitated a physical search more akin to foraging. Having to conduct physical search increases the opportunity to make looks to task irrelevant objects and also requires more looks to room features, such as cupboards, including those not containing target objects in order to forage for objects hidden from view. The substantial reduction in looks to task irrelevant objects suggests that as participants learned their environment these costs associated with foraging in a new environment were overcome and participants dedicated more of their gaze to task relevant features and objects. We know from the literature on foraging that people tend not to revisit previously searched locations, (Gilchrist et al., 2001; Smith et al., 2008) and as a result of this, cupboards not containing targets would have been unlikely to have been re-searched, therefore limiting visual exposure to the contained distractor objects and reducing the number of looks to task irrelevant objects.

6.4.6 Active search as foraging and the effect on eye-hand latencies

Foraging could be viewed as an instance where action aids vision, however the primary focus of this thesis is to examine vision for action. The well-known tea making study conducted by Land et al. (1999) demonstrated that vision is proactive, with fixations preceding motor manipulations by approximately half a second. The results from the present study support this, with the eye leading the hand by just over half a second, this lead by the eye occurred in over ninety-five percent of all of the actions. The eyes have been shown not only to lead the hand but also other types of action, for example the foot when walking (Land & Tatler, 2009) or a cursor hitting a target during a visuomotor task (Sailer et al., 2005). We know that there are occasions when the latency with which the eye leads the action can change depending on circumstances like task experience, for example, more experienced drivers tend to look further into the apex than inexperienced drivers who essentially keep their fixations close to the current action rather than visually leading or anticipating the next one (Land & Hughes, in Land, 2006), similarly Land and Tatler (2009) describe how in a study conducted by Land and colleagues, they examined where people look when walking up stairs and demonstrated that after initial learning of foot placement on stairs, the latency between fixation and foot placement lengthened. As experience is accumulated there appears to be a tendency toward longer fixation lead times: Sailer et al. (2005) demonstrated that as individuals learn to co-ordinate eye and hand movements, during the early stages of learning of a novel visuomotor task gaze typically pursues or stays close to the cursor, then as the task becomes learned fixations begin to predict desirable curser locations until finally the task is learned and the skill is being refined and gaze is directed on the target, thus increasing the latency between fixation and action. Visual experience with an object may also impact eye-hand latencies, in that, during a task we are typically able to look as many times to objects in the room as we desire, in the Mennie et al. (2007) study it was demonstrated that gaining visual experience alone in the form of look-ahead fixations to an object can impact the subsequent eye-hand latency with that object. Therefore we know that not only do the eyes lead the hand during actions but factors such as experience with objects or tasks may change the timing of this latency.

6.4.7 Eye-hand latencies affected by experience

In line with the present studies finding that eye-hand latencies change as familiarity is acquired Foerster, Carbone, Koesling, & Schneider (2011) found that across a 14 day training period of learning a cup stacking task eye-hand latencies decreased; however when the decrease in trial duration was accounted for, and the latencies were expressed as a proportion of task time, the proportion increased. Foerster et al. speculated three possible reasons as to why this might have occurred, firstly they suggested there may be a biological limit to eye-hand spans and that the cognitive processing between visual input and motor output had gotten as short as biologically possible resulting in eye-hand time spans ceasing to decrease whilst the performance speed continued to accelerate. Secondly, they proposed that it may be due to a slower decrease in eye-hand time spans than trial durations or thirdly the dynamics of the coordination of the eye-hand spans may have improved and resulted in a faster performance driving fixations to the next object while the current manipulation is still being completed.

Considering the literature demonstrating an increase in eye-hand latency time with experience (Land & Tatler, 2009; Mennie et al., 2007; Sailer et al., 2005), it is perhaps somewhat surprising that the present study found that the latencies significantly decrease as experience is accumulated. Furthermore, given increases in time spent looking ahead in the present study, the finding that there was no increase in eye-hand latencies as experience is acquired is somewhat surprising. Perhaps here it is important to consider the distinct elements of the present study that may have contributed to this. Firstly, the physical environment that the study was set in meant that not all of the target objects were visible without requiring some physical searching (akin to foraging). This is different to all of the other previously discussed studies and may have restricted the opportunities for participants to direct their eyes to the next object much before they were able to initiate the related motor act. For example, if a person was pouring tea into a cup placed on the worktop on one side of the room, and the next action was to fetch the milk from the fridge in on the opposite wall, this would require a 180° turn, locomoting across the kitchen and then opening the fridge before the milk could even be fixated. In this case the opportunity to fixate much before the initiation of the reach for the object is restricted somewhat. The layout of the environment and the placement of objects may have prohibited acceleration in tempo of eye-hand latencies as familiarity was acquired, however while this may account for not finding an increase in eye-hand latencies it does not easily explain why we found a decrease.

Secondly, the style of task we used was a sequential task, which required the completion of many small sub-goals in order to achieve the main task goal of making tea. Each small component of the task was distinct from others, with the exceptions of picking up/putting down objects and opening/closing cupboard doors. This is different to all of the other tasks used in other studies which have tended to use simple tasks with repeated actions, for example picking up and setting down blocks (Mennie et al., 2007), or walking up stairs (Land & Tatler, 2009), and do find a shortening latencies. This may be because after completing the same act a few times, actions may have become somewhat automated which in turn may have freed up visual recourses sooner, since continuous visual guidance was not necessary. Maximising opportunity to fixate on the next object to be manipulated while the hands were still engaged in the previous manipulation would therefore produce longer eye-hand latencies. In the present study, whilst the steps of the task as a whole can be somewhat automated (Land et al., 1999) many of the sub-goals require complex manipulations of objects, for example the teapot lid has two internal notches which must be aligned with the corresponding notches in the teapot, therefore visual guidance is required at least until the lid is aligned. Therefore, opportunity to fixate the next object to be used is reduced and is likely to be temporally much closer to the point where the next motor action also begins.

Throughout the task there are many examples of sub goals that would require visual guidance for prolonged periods thus limiting instances where the fixation to the next object could occur earlier.

Thirdly, if it is the case that during the initial part of a fixation to an object about to be manipulated, visual cues about the size, density and weight of an object are being processed (Cole, 2008) rather than object specific information for everyday objects being retrieved from memory (Wolpert, Ghahramani, & Jordan, 1995) then perhaps, the eye-hand latency reflects this time needed to process the features of the object that are required to be known for action. Cole (2008) suggests that even with familiar objects, we visually process the size and combine that knowledge with prior knowledge or expectations about the density of the object. Thus, although familiarity does not rule out the need to perform an on-line visual analysis altogether, accumulating experience with these objects may reduce the need for this processing or may make the processing become faster therefore inducing shorter the shorter eye-hand latencies we demonstrated in the present study.

6.4.8 The effect of object properties on eye-hand latencies

Eye hand latencies can be affected by the accumulation of experience with the objects in terms of size, weight and density (Flanagan & Wing, 1997; Randall Flanagan, King, Wolpert, & Johansson, 2001.; Wolpert, Ghahramani, & Jordan, 1995). Cole (2008) demonstrated that even with familiar objects, on-line visual assessments are completed in order to determine the appropriate acceleration of movement and size-related finger tip force rather than rely on memory. Whilst we utilise priors regarding the likely properties of objects it would seem that these are quickly updated with visual information in order to best guide interactions with the object. In the present study there were no novel types of objects used, however the exact objects had, to our knowledge, never been seen prior to day 1 of testing. Of the target objects interacted with in our study, the longest eye-hand latencies were those where the kettle was the target object, compared to the shortest latencies being for both the milk lid and the tea caddy lid. There was no significant change in latencies by object type across days. To speculate then as to why there was a difference in latencies depending on the object type, we must consider what it is that is different about the objects themselves. The object with the longest latencies was the kettle which also happens to be the largest object used in the task, whereas the shortest were for lids, the smallest objects, thus the differences in eye-hand latencies may reflect something about the size of the object, perhaps the larger the object, the longer the eye-hand latency. If this was the case then we could order the objects used by size and predict the increase in associated eye-hand latencies accordingly, however, if we consider the example of three objects used in the task and order them by size from largest to smallest: kettle, tea caddy and teaspoon, the eye hand latencies should increase in the same pattern, with the largest latency for kettle and the shortest for teaspoon, however this is not what happens. The results demonstrate that the largest latency is for kettle but the shortest latency is for tea caddy rather than spoon, therefore the size of object was unlikely to be the sole cause of differences in eye hand latencies.

Along with processing features such as height and weight of an object during the eye-hand latency, it is conceivable that some processing of other object characteristics also occur, such as when the kettle is boiled it is hot to touch, which could contribute to longer eye-hand latencies for some objects over others. So risks associated with certain objects may influence our visual behaviour when interacting with them. We know from recent work that the way we intend to interact with an object influences not only where we fixate on an object, but also the reaction time to perform the action. For example, Belardinelli, Herbort, & Butz (2015) measured fixations and RTs for both a touch screen and pantomime gestural task with three task conditions where participants saw real everyday objects on a screen and had to either classify something about the object (whether it could hold liquid or not – passive condition), or to lift or open the object (active conditions). The results showed that RTs were significantly shorter for the passive task than for the active tasks, and that the first fixation across all three tasks was not distinguishable between conditions but by the second fixation it was possible to predict the action to be undertaken (in this case lifting or opening) based on the placement of the second fixation. The longer reaction times associated with having to be active with an object as found by Belardinelli et al. (2015) may suggest that extra processing and/or planning has to occur during this time. Thus having to be active with an object that may pose a risk might influence how long we fixate on the object before initiating action. The present study finds that the kettle elicited the longest eye-hand latencies, whereas objects such as lids and teaspoon elicited the shortest, the kettle could be considered to be an object associated with a high degree of risk (particularly in its boiled state), whereas the lids, pose no significant risk. Cinelli, Patla, & Allard (2009) demonstrated that threat to stability (when locomoting through oscillating doors) changes visual behaviour and elicits more “online” control to directly guide behaviour. Therefore motivation to drive the eyes to the next target object may be influenced by the risk level associated with the objects, since it would be beneficial to allow for extra processing time and planning for objects that would pose risk when executing an action upon them. Ballard and Hayhoe (2009) comment on the difference in the way participants dealt with a knife spreading jam compared to the same knife spreading peanut butter in the sandwich-making task conducted by Hayhoe et al. (2000). Ballard and Hayhoe noted that participants tended to fixate more on the knife with jam whilst guiding it to the bread than the one with peanut butter, they concluded that this was due to the viscosity of the spreads with jam being considered more ‘precarious’ and requiring extra visual guidance.

6.4.9 The order of sub-goal completion changes as familiarity is acquired

As familiarity with an environment and its objects increases, the order actions are completed in become more similar, suggesting that familiarity affords the ability to plan consecutive actions and specify an action script. Norman and Shallice (1986) proposed that as we learn a task we produce scripts and run those from memory resulting in less on-line processing, similarly Land, Mennie, and Rusted (1999) referred to well practiced task performed with little conscious involvement as automated tasks. Most everyday tasks necessitate that actions are executed in an order that leads to the overall achievement of the main task goal. In real life the extent to which the order of the task is fixed or flexible varies. In the present study, the order the task sub-goals had to be completed in combined some fixed elements of the task which had to be completed before others, for example the kettle had to be boiled before it could be poured into the teapot, but also some flexible elements, for example the milk could have been added to the teacup before the tea or vice versa. Deciding which order to complete these sub goals in could either be planned ahead or decided in a more ad hoc, moment-to-moment basis. Our finding that the order changes and becomes more similar across days indicates that either there is a two-stage approach occurring depending on familiarity with the environment, initially an ad hoc system of deciding which sub goal to complete next may be employed, progressing to planned, or that all of the days are planned but that the plan simply changes across days. If the two-stage approach of firstly ordering sub-goals in an ad hoc manner occurs this may be because in a novel environment there would be no prior knowledge of the exact locations of objects, therefore, planning an action script in advance based on the task without taking the environment into account could potentially be costly in terms of time and effort, with more revisiting locations, errors and searches required. In a familiar environment it would be more advantageous to utilise prior knowledge to develop an action script based on the specific layout of the environment and the locations of the objects.

Humphreys, and Forde (1998) examined activities of daily living in both patients with impaired everyday-life behaviour and in normal participants and had normal participants list the actions they would usually carry out to complete a task. They found that single component actions could be grouped into subroutines which themselves could be grouped into larger routines, and so on, see Figure 3 below.

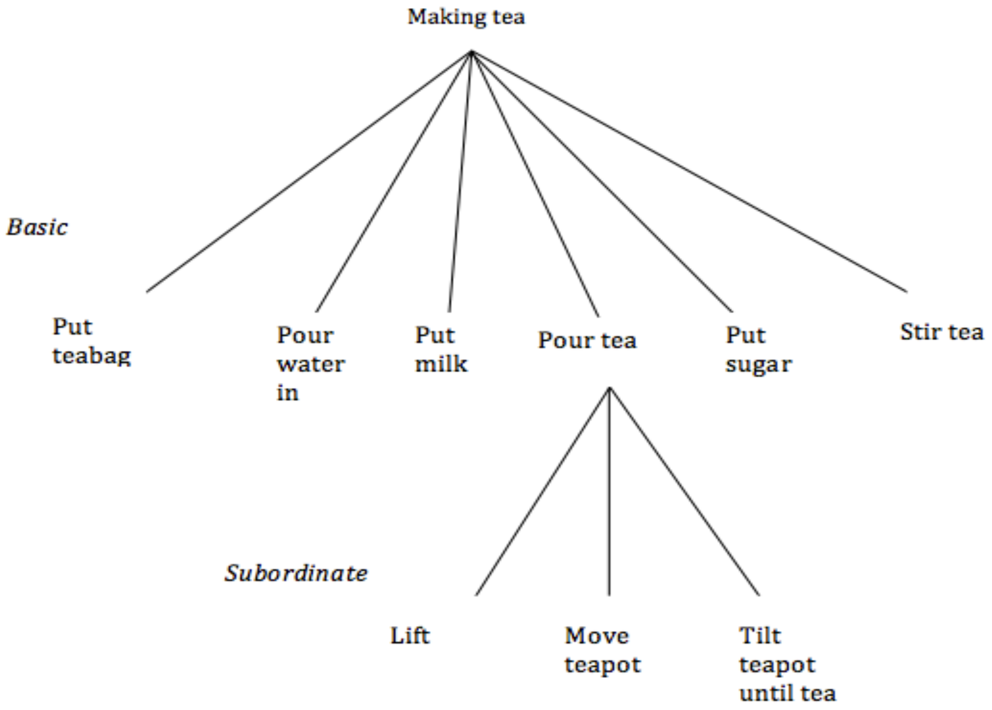


Figure 6.15. Reproduced from Humphreys & Forde (1998)

Humphreys and Forde argued that the processing system is arranged in hierarchical layers related to discreet levels of task structure and that simple actions involve the coordination of multiple schemas associated with different levels of temporal structure. However, what this description does not do is describe how decisions are made regarding which sub-goal to complete next when there is a choice of appropriate next actions. Cooper and Shallice (2000) put forward a model where a network of hierarchically organised schemas process competition of activation for the selection of routine actions. They argue that schemas are partially ordered methods for achieving goals and that the most appropriate schema for any situation depends on several factors, including the objects available in the environment and individual preferences. The activation of a schema resulting in an action varies over time, the authors identify four sources that influence the activation, namely, the presence of objects, ‘top-down’ influences from higher level schemas (see Figure 6.16 below for an example of the hierarchical structure), a lateral influence whereby the system ensures competition between schemas inhibit the non-selected action and self influence which partially counters the lateral influence. Our results may further this idea and suggest that the weighting of these four sources may depend on the level of our prior knowledge of these objects and their locations in the environment, with no or limited prior knowledge weighted to respond more to the presence of objects and the lateral influence to ensure that a choice is made if more than one object is present, and then as familiarity is acquired, stronger influence of “top-down” schemas and self influence for the activation of schema.

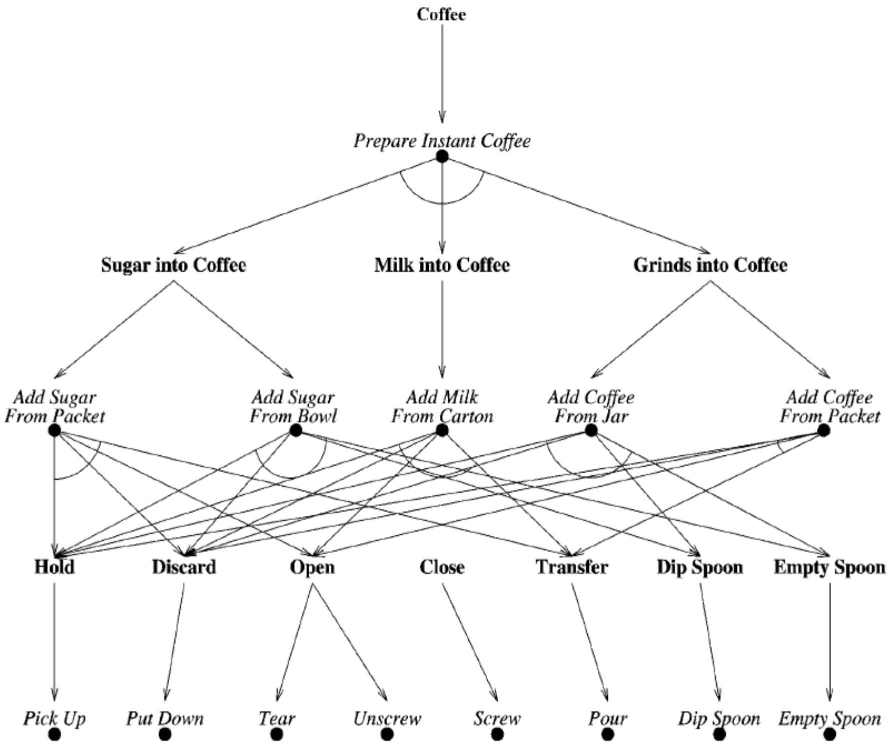


Figure 6.16. Schema/goal organisation for the overall task of preparing coffee. Schemas are indicated in italic and goals in bold type. Reproduced from Cooper & Shallice (1998)

6.4.10 Decision-making and familiarity acquisition

The other possibility regarding the tightening of action scripts is that it is not related to planning a full script per se, rather experience with the environment and its objects simply improves the decision making as to what to do next at each stage of the task. Thus we may still be working moment-to-moment but are simply better at it, faster and with less effort expenditure. We know that in sport, variation and experimentation with skills, techniques and tactics affords greater opportunity to practice the skills necessary for anticipation and decision making (Ward & Williams, 2003; Williams et al., 2011). For example, Vaeyens, Lenoir, Williams, Mazyn, and Philippaerts (2007) demonstrated that experienced soccer players demonstrated superior decision making skills, were faster and more accurate and typically made their decisions based on anticipation rather than reacting to the events. Experience changes the speed and way we make decisions, therefore this may account for the increase in similarity of action scripts across days found by the present study.

6.4.11 Look-ahead fixations and planning

Look-ahead fixations may reveal something about the planning of actions yet to be completed. More LAFs are made as familiarity is acquired suggesting that we may better able to plan actions further in advance when we are familiar with the environment and its objects. Pelz and Canosa (2001) noted that during an active task, in their case hand washing, a small number of fixations were made to objects that would be relevant for actions in the near future (look-ahead fixations). According to Pelz and Canossa, one of the purposes of these LAFs is to facilitate our ability to task switch, in a sequential task like tea-making the sub goals are often quite distinct and thus could be considered as micro tasks which we have to switch between. The results of the present study indicate that across the 10 task days, in relative terms there was a significant increase in the proportion of time spent making look-ahead fixations. If we consider the argument made by Pelz and Canosa (2001) if the purpose is to support task switches, the same number of core sub-task elements have to be completed on day 10 as on day 1, in fact the increase in similarity of order of task completion across days suggests that this element of the task (i.e. switching between sub-tasks*)* should have become simpler, thus we would have expected to see a reduction in LAF behaviour.

Mennie, Hayhoe, and Sullivan (2007) argue that look-ahead fixations are purposeful and play a role in planning. Their results indicated that the eye-hand latencies to target objects that had previously been the subject of a look-ahead increased by 122 ms and were accompanied by a more accurate reach. However, the present study found the opposite direction of proportion of time spent making look-ahead fixations and a reduction in eye hand latencies. One of the reasons for the conflicting finding may be that the task that participants completed. In the study conducted by Mennie et al. the actions were scripted, in that participants were instructed what to do at each stage of the task, for example to reach out piece one and join to piece two, so decision making for the individual was minimised, whereas, the present study used a sequential task where some of the sub-task elements can be completed in a flexible in the sequential order. Thus planning the next action to be made would require some form of decision making to inform the choice. Look-ahead fixations may represent something about planning the next act in the sequence but it is also possible that another type of planning is occurring, in order to achieve many of the sub-goals in the task of tea making in a typical kitchen, many distinct motor actions are required for each task element, furthermore whole scale body movements and even locomotion is necessary, therefore it may be that LAFs represent information searches which facilitate motor planning (Hayhoe, Shrivastava, Mruczek, & Pelz, 2003a).

6.4.12 Conclusion

The present study set out to explore the influence of acquiring of familiarity with an environment and it's objects, both in terms of general and visual behaviour. We found that as familiarity is acquired, the overall task completion time significantly reduces and the composition of time spent performing specific behaviours changes. By the last day of the task, significantly more time was spent performing look-ahead fixations and significantly less time visually exploring the environment and its objects, Due to the layout of the objects in the present study, i.e. not all were visually available without physical search, the task was more akin to a foraging task rather than a traditional visual search. In this light our results correspond with the literature on visual behaviour during foraging tasks (Gilchrist et al., 2001; Smith et al., 2008) and may suggest that the reduction in time spent visually exploring the room reflects an increase in the readiness to rely on memory and a higher motivation to avoid having to revisit locations which previously had not contained the target object. The order in which sub-goals of the task were completed also changed across days with more similarity towards the end of the task, suggesting that even although the task of making tea is well learned, performing the task in a new environment with new objects can impact the order in which we perform the task itself. We know that vision is proactive and the eyes lead the hand during an active task, and our results are in line with previous work demonstrating this (Hayhoe et al., 2003; Hayhoe, 2000; Land et al., 1999; Land & Hayhoe, 2001) however the finding that eye-hand latencies get shorter as familiarity is acquired is a relatively novel finding, suggesting that this lead time is not fixed. We are still unsure exactly what the purpose of eye-hand latencies are but if some level of processing information about the object and its properties is occurring, we can speculate that as our representations strengthen with experience, we require less time to perform on-line processing before making physical contact with the object. The subsequent elements of physical manipulations on objects after the arrival of the eye also get shorter as familiarity is acquired which may indicate that even for familiar object types the *power law of practice* may influence our behaviour and speed up our performance. Ultimately we can conclude that behaviour, both general and visual changes as familiarity with an environment and it's objects is acquired and therefore future studies examining eye movements in natural environments should pay attention to the level of familiarity with not only the task but also the objects and environment the task is conducted in. We cannot tell from the present study what is being represented during the acquisition of familiarity, whether it is only information about the task relevant objects, the spatial layout of the environment or even task irrelevant objects present in the environment, or a combination of these factors. The next chapter will examine the effect on visual behaviour of having developed familiarity for an environment and objects required to make tea by switching task after the initial 10 familiarity acquisition days.

Chapter 7 What a task switch reveals about incidental encoding during the acquisition of familiarity

7.1 Introduction

The previous chapter demonstrated how eye movements change during the acquisition of familiarity of both an environment and the objects contained within the environment. As participants became more familiar with the environment various aspects of their visual control and behaviour changed. Participants were faster at completing the task and the order in which they performed the sub-goals of the task became more similar across days. As more experience was accumulated with the environment, people looked less at task irrelevant objects, spent less time visually exploring the room and displayed shorter eye hand latencies. These types of changes are consistent with an increasing use of representations of the environment and objects.

During the acquisition of familiarity clearly some information about the environment and the objects is being encoded but we cannot tell from the previous chapter what these representations include. It may be that the visual system represents only the information necessary for the task (Hayhoe, 2000) or that the representation is more inclusive and incorporates other objects and features of the environment that are encountered incidentally during viewing (Henderson & Hollingworth, 2003a, 2003b; Hollingworth & Henderson, 2006; Hollingworth, Williams, & Henderson, 2001; Williams, Henderson, & Zacks, 2005). The kitchen used both in the present study and in the experiment described in Chapter 6 included many other kitchen relevant objects, present for the entire acquisition phase. These objects need not have been encoded in order to complete the objective of making tea, however there were fixations made to these objects and so the present chapter will investigate whether these task irrelevant objects were encoded and represented, despite not being necessary, by way of a task switch at the end of the initial familiarity acquisition period allowing us to test the question of whether objects and environment are incidentally encoded during the acquisition of familiarity.

Whenever an active task is performed in an environment we have prior knowledge of, we can either use memory of the layout of the room and objects to guide our vision and behaviour or we can simply sample when necessary the environment itself by moving our eyes (or a combination of both). The idea that visual representations are both local and transient was put forward by several authors who argued that highly detailed visual representation is limited almost entirely to the currently attended object (O'Regan, Rensink, & Clark, 1999; O'Regan, 1992; Rensink, 2000, 2002; Simons & Levin, 1997; Wolfe, 1999) and that only a few specific details of the objects themselves are encoded (Chun, 2003; Simons & Levin, 1997; Wolfe, 1999). This view also asserts that representations of objects are severely impoverished whenever gaze is not directed at said object, however, there is considerable evidence that this is not the case and actually long term memory for scenes is quite accurate. Observers are able to recognise thousands of pictures presented earlier in a study (Nickerson, 1965; Shepard, 1967; Standing, 1973; Standing, Conezio, & Haber, 2013), can effectively distinguish between objects that appeared in a studied scene and conceptually related but visually dissimilar distractors

(Friedman, 1979; Hollingworth & Henderson, 2002) and can detect changes to previously fixated objects even when the change occurred well after the fixation (Hollingworth et al., 2001). Originally it was assumed that the level of detail that is encoded for complex scenes is limited and that although people are able to distinguish large sets of old pictures from new distractor pictures, their ability to detect missing elaborative visual details is more limited (Pezdek et al., 1988), however a number of recent studies have challenged these assumptions and demonstrated that visual memory representations often contain considerable detail (Brady, Konkle, Alvarez, & Oliva, 2008; Hollingworth, 2006; Mitroff, Simons, & Levin, 2004; Simons & Rensink, 2005).

In order to acquire familiarity with an environment and its objects, information about the geometry of the room (Sturz, Gaskin, & Roberts, 2014) and the global layout of the room features would have to be encoded along with more local information such as the positions of objects in the room. Similarly it would be essential to encode something about the objects themselves, like the layout, appearance and associated functional properties. The extent to which this occurs could vary depending on a number of factors. The task relevance of objects may influence the overall detail of representation accumulated for an environment, it may be that only task relevant objects are represented (Hayhoe, 2000) or it could be representations are not bound to task as is suggested by the findings that even for task-irrelevant objects simply looking at an object does indeed result in memory retention (Castelhano & Henderson, 2005; Hollingworth, 2006; Williams, Henderson, & Zacks, 2005).

The finding that participants incidentally generated memory for non-target objects during search tasks and such memories facilitated performance on subsequent searches (Hout & Goldinger, 2010) suggests that detailed visual information is encoded incidentally. Võ & Wolfe, (2012) found that search was only facilitated when the target had previously been searched for rather than simply fixated during a memorization task, suggesting that task matters for what is encoded for later searches. However, Hollingworth, (2012a) found that if an object was fixated it did indeed facilitate subsequent search performance irrespective of the initial task, implying that representations are more detailed and less task bound. The influence of task instruction on memory representations for objects was also examined by Tatler and Tatler, (2013) in a more ecologically valid paradigm by having participants view a number of real objects in a room under three viewing conditions, participants either viewed the objects with no task instruction and no indication that there would subsequently be a memory test, with the instruction to remember as much as they could about all of the objects in the room and finally a condition whereby participants were instructed to remember only certain objects in the room. Although task instruction did modulate performance, the authors found that participant performed above chance for all question types even in the ‘free-viewing’ condition where participants had to expectation of a subsequent memory test, which confirms that memory encoding includes incidental encoding of objects not crucial to the task.

In certain circumstances incidental encoding for non-target objects occurs however, the extent to which this happens over an extended period of time during active sequential tasks has yet to be examined. Since we know that task can modulate both subsequent visual behaviour and which objects are encoded in memory, we might expect that the demands of the task would promote more use of memory and therefore even task irrelevant objects would be encoded during the initial familiarity acquisition phase. Alternatively it may be the case that because completing sequential tasks with multiple sub-goals already presents a heavy processing load, storing any more details than needed for the task in hand would add more load to an already occupied set of systems. If this is the case we would expect that only the task relevant objects would be encoded for representation.

Traditionally experimental paradigms test whether task irrelevant objects are encoded incidentally in explicit memory tests, either by asking viewers about objects that were present but not previous visual targets (Tatler & Tatler, 2013), or with the use of alternative-forced-choice questions about the presence and properties of objects (Monica Castelhano & Henderson, 2005). Alternatively, the memory test may present a previously viewed scene with the viewer asked to search for a formerly task irrelevant object (Võ & Wolfe, 2012), the time to fixate the new target is then measured. Here we extend the latter approach by having participants act on objects that had been present during previous trials but had been irrelevant to the initial task. Although we cannot measure search time to new target directly, since the order sub-goals are performed is self directed by the participant we would expect that if non-target objects are encoded incidentally then this should be apparent in their visual behaviour. The extent to which participants would have to essentially start anew in a familiar environment if the task now requires use previously non-target objects should provide some insight as to whether task irrelevant objects are incidentally encoded during the acquisition of familiarity.

In the previous chapter participants made tea in the kitchen where other objects not needed for that task were also present. The non-target objects were carefully selected to comprise two sets of objects required for two other tasks. We know from the previous study that on day 1 looks to task relevant and irrelevant objects was roughly even, whereas by day 10, only 20% of the looks made were to task irrelevant objects, however we do not know whether these objects were incidentally encoded. In order to investigate this the same cohort of participants who completed made tea for 10 days in Chapter 6 were asked to attend for a further two days being led to believe that they would be completing the same task (making tea) for all twelve days. However on the eleventh day, participants were informed immediately before entering the room that they were instead required to perform a different task (either making a sandwich or cup of coffee), this was repeated on day twelve, where participants were asked to complete a third task instead (either making a sandwich or a cup of coffee, whichever they had not completed on day 11). The alternate tasks used objects that had been present in the environment for the entire previous 10 days, and in the present study we are interested in examining whether incidental encoding occurred for these ever present objects, despite them being task irrelevant for the initial 10 days. Whether task irrelevant objects are incidentally encoded when learning a new environment should be apparent in terms of how participants treat the visual guidance of dealing with task switch objects. The way in which that would manifest in terms of the effect on behaviour and eye movements with the new task objects is the principle aim of this chapter. Based on the findings of the previous chapter, if the sandwich and coffee making objects had been incidentally encoded we might expect that the visual behaviour would look more like that found for the end of the familiarity acquisition phase rather than on day 1. Specifically, we would predict that the proportion of time spent visually exploring the environment would be shorter than on day 1 but not different from day 10, and that there would be a similar proportion of time spent making look-ahead fixations and looks to task irrelevant objects as on day 10. Anticipating the effect of incidentally encoding coffee and sandwich objects during the ten tea making days is more difficult to predict, it may be that if these objects were incidentally encoded then the eye-hand latencies would look like the short latencies demonstrated on day 10 of tea making, however since the objects had not previously been physically interacted with it may be that the category and location was incidentally encoded but that there would be no change in eye-hand latencies since no motor familiarity was established.

7.2 Method

7.2.1 Participants

The same 10 individuals who participated in the study presented in Chapter 6 participated in the experiment for the current chapter. After making the in the same kitchen for 10 days, on the subsequent two days the same participants attended a recording session where, unbeknownst to them previously half of the participants were required to make a sandwich and half to make a cup of coffee, then on day 12 to make whichever (coffee or sandwich) that they had not made the day before.

7.2.2 Materials

The same kitchen used in the preceding chapter was used for this study with all of the previous target objects still present as distractor objects in this experiment. There were two sets of target objects (one for each task carried out on the two consecutive days), one set contained coffee making equipment and the other contained sandwich making equipment. All of the objects for both tasks had been present for the familiarity acquisition study presented in the previous chapter. Note deliberate care was taken to ensure that the coffee making task required different objects from the tea making task, instead of the kettle coffee was made using a percolator, a mug replaced the teacup and cream instead of milk was added to the coffee. Participants were told which objects to use but nothing about their location or the order of sub goal completion. The number of objects for both task switch days were balanced to be equal to the number of objects used for the tea-making task in the previous chapter. All of the task switch objects were placed around the room in similar places to the tea making objects so were visually available for the 10 tea making days in the previous chapter, on the task switch days these objects were initially presented in the same locations they had occupied for the previous 10 days.

7.2.3 Design

A within subjects design was used with all participants completing both the tea making task and the sandwich making task, the order of the two tasks was counterbalances across participants.

7.2.4 Procedure

All participants were under the impression that they would be completing the same task for 12 consecutive days. On day 11 participants were informed just before entering the kitchen that actually the task had changed and they were either required to make either a cup of coffee or a peanut butter and jam sandwich. For participants in the coffee first condition, they were asked to make a cup of fresh coffee in the mug, using the coffee machine and to add cream and sugar, whilst participants making the sandwich were asked to make a peanut butter and jam sandwich, using the white chopping board and a plate to serve, the extra steps in each task were to ensure that an equal number of objects were used on day 11 and 12 as had been used to make tea for the previous 10 days. On day 12 the conditions were counterbalanced so that the participants who made coffee on the previous day instead made a sandwich and vice versa.

7.2.5 Analysis

We used two series of analysis to examine how incidental encoding during tea making affected visuomotor behaviour on coffee and sandwich making. In the first section, we look at general effects of having been in the room for 10 days. In the second section, we ask whether specific fixation behaviour on items during tea making that were irrelevant for that task affects visuomotor behaviour when the items now become relevant. For example, a participant may have fixated the coffee machine when making tea, but as it was not relevant we do not know what, if anything, was encoded about that object.

In the first section of analyses we use LMMs to examine whether vision and action in coffee and sandwich making more closely resembles day 1 (when the room was entirely novel) or day 10 (when the room was familiar). To do this we used a dummy coded factor that made separate comparisons between a measure on day 11 or 12 to day 1, and to day 10. Participant was included in the model as a random effect.

In the second section, we used ANOVAs and LMMs to examine whether visual behaviour on objects was related to the amount of time spent looking at those objects on day 1-10 of tea making. This allowed us to ask whether information about the objects that would influence how participants acted on them was encoded incidentally when fixating it as an object that was irrelevant to their current task (making tea). When each participant was only represented once (for example, the time it took them to complete the task), we used ANOVAs, but when each participant had multiple measures (such as eye hand latencies), we used LMM models with both participant and object included as random effects.

For these analyses, we did not include P10, as they did not have data from day 1 and it was therefore not possible to know if they looked at the coffee and sandwich objects on that day.

7.3 Results

7.3.1 Effect of spending time in the room

7.3.1.1 Exploration time

An LMM was used to examine whether the raw amount of time exploring the environment was more similar to day 1 of tea making (when participants were *unfamiliar* with the environment) or day 10 of tea making (when participants were *familiar* with the environment). The results are shown in Figure 7.1. Participants spent significantly less time exploring the room in both the coffee ($\beta = 37.498$, $SE = 8.083$, $t = 4.639$, $p < .001$) and sandwich making ($\beta = 33.762$, $SE = 7.712$, $t = 4.378$, $p < .001$) tasks when compared to tea making on day 1. However, exploration times were not significantly different from exploration times in on day 10 of tea making in coffee ($\beta = -4.234$, $SE = 8.083$, $t = -0.524$, $p = 0.891$) or sandwich ($\beta = -7.969$, $SE = 7.712$, $t = -1.033$, $p = 0.554$) making.

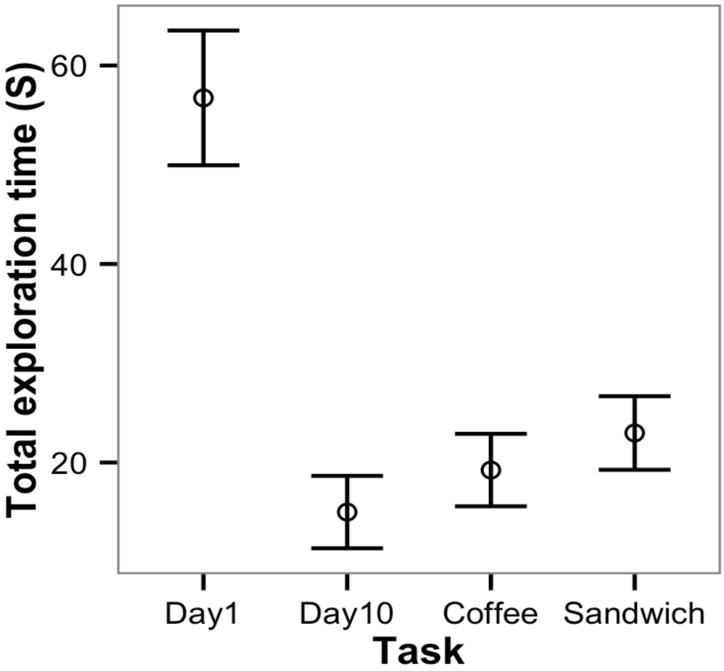


Figure 7.1 Mean and SE of the total exploration time in the different task conditions. Means presented here are adjusted to remove the between-subject variability using the elegant method of Cousineau (2005): plotted value = observed value – subject mean + grand mean, although the raw data were used in the statistical analysis.

7.3.1.2 Task irrelevant looks

Across days 1-10 of tea making, the proportion of looks to task irrelevant objects decreased. Here, we query how the proportion of looks aimed at task irrelevant objects changes when participants were asked to complete a novel task in a now-familiar environment. We measured the proportion of looks of task irrelevant objects to task relevant objects. We then used a LMM to compare these proportions on the coffee and sandwich tasks to day 1 and day 10 of tea-making, with participant included in the model as a random effect. As with the proportion of time spent exploring the environment, participants aimed significantly less of their looks to task irrelevant objects compared to day 1 of tea making in both coffee (*β* = 0.399, *SE* = 0.043, *t* = 9.325) and sandwich making (*β* = 0.253, *SE* = 0.039, *t* = 6.518), but were not significantly different from day 10 (coffee: *t* = 1.868, sandwich: *t* = -1.709). A final model compared the task irrelevant looks in coffee to sandwich making, revealing that significantly more looks were made to task irrelevant objects when making sandwiches (*β* = 0.146, *SE* = 0.044, *t* = 3.339, *p* < .001).

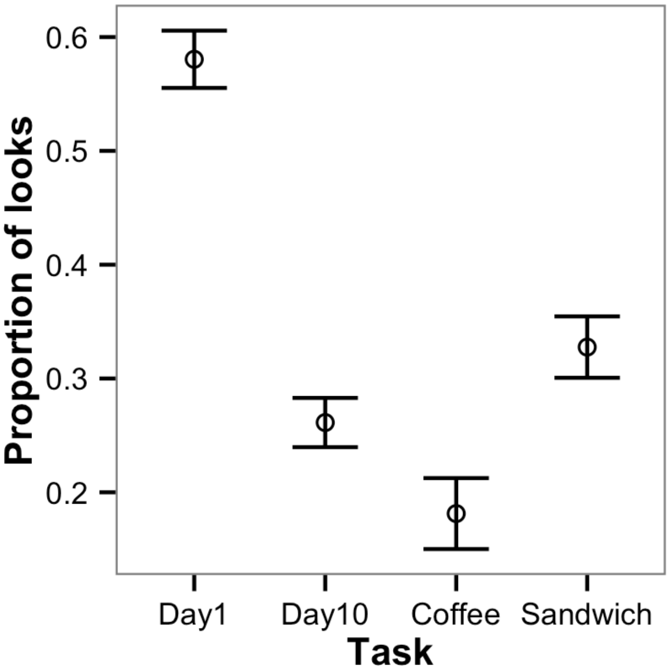


Figure 7.2 Mean proportion of looks to task irrelevant objects with between-subject variability adjusted SE's.

One possible explanation of the reduction in proportion of time spent visually exploring the environment and looking at task irrelevant objects is that having encoded where the tea objects are, after a task switch participants need not fixate the tea objects thus the reduction in these looks in turn reduces the time spent visually exploring and looking at task irrelevant objects.

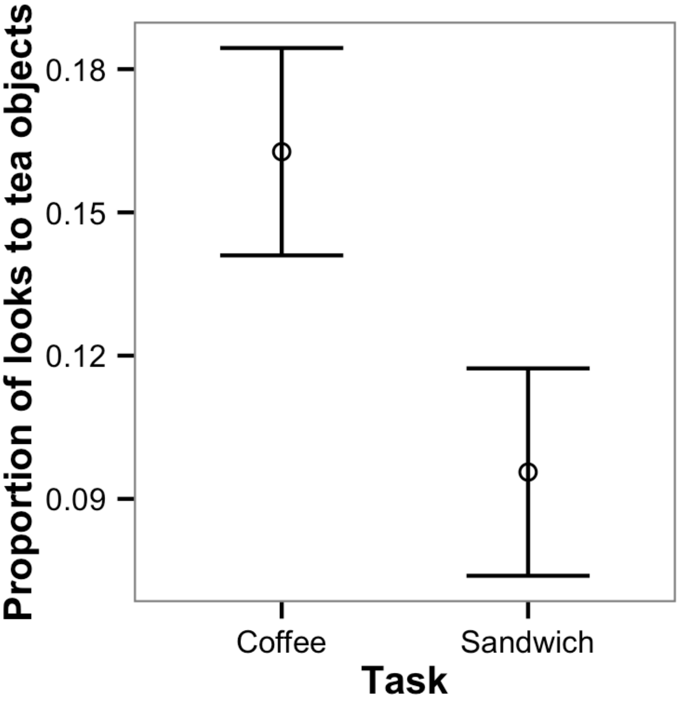


Figure 7.3 Mean proportion of looks to tea making objects (task irrelevant objects) during coffee and sandwich making.

7.3.2 Effects of looks to object during tea making

7.3.2.1 Task completion time

We looked at the relationship between the total frequency of looks to objects on day 1-10 of tea making, and the time it took participants to complete the task using an ANOVA. We included task (coffee or tea) and the total frequency of looks as independent variables, with an Error term indicating the within-group nature of the IVs.

There was a significant positive relationship between the total frequency of looks to tea and coffee items and the time it took to complete the task on days 11 and 12 (F (1,7) = 36.99, p < 0.001; Figure 7.4). There was also an effect of task, with coffee taking longer to make than sandwiches (*F* (1,7) = 46.47, *p* < .001). These effects were the same when we looked at the duration of these looks (both p values < .001; Figure 7.5).

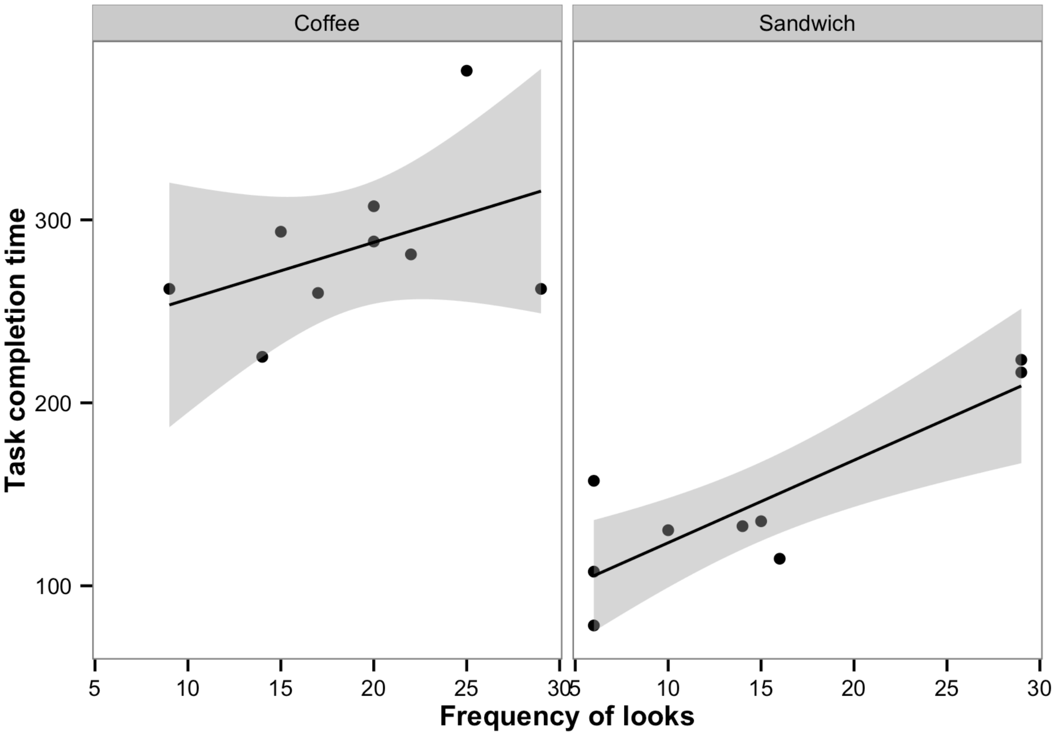


Figure 7.4 Frequency of looks to the incidental objects across the 10 days of tea making and subsequent task completion time for both of the incidental task switch days (coffee left panel, sandwich right panel).

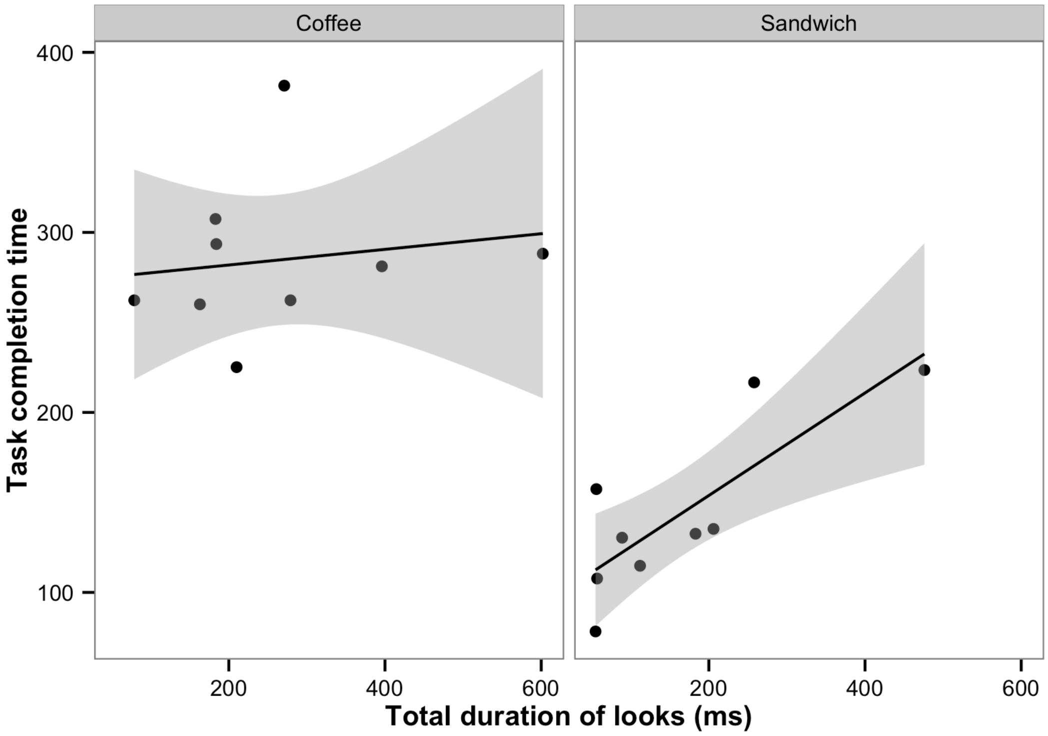


Figure 7.5 Duration of looks to the incidental objects across the 10 days of tea making and subsequent task completion time for both of the incidental task switch days (coffee left panel, sandwich right panel).

This is a somewhat surprising result, as we might expect that if participants looked at coffee or sandwich objects that they might encode information that would allow them to complete the task more rapidly. One possible cause of this effect is that some participants imposed a slower tempo on tasks in general. To check this, we examined whether the median time to complete the tea-making task across the 10 days could predict the time it took to make coffee and sandwich. There was a significant relationship between the two factors, with people who took longer to make tea similarly taking longer to make coffee and sandwiches (*F* (1,7) = 6.163, *p* = 0.042; Figure 7.6).

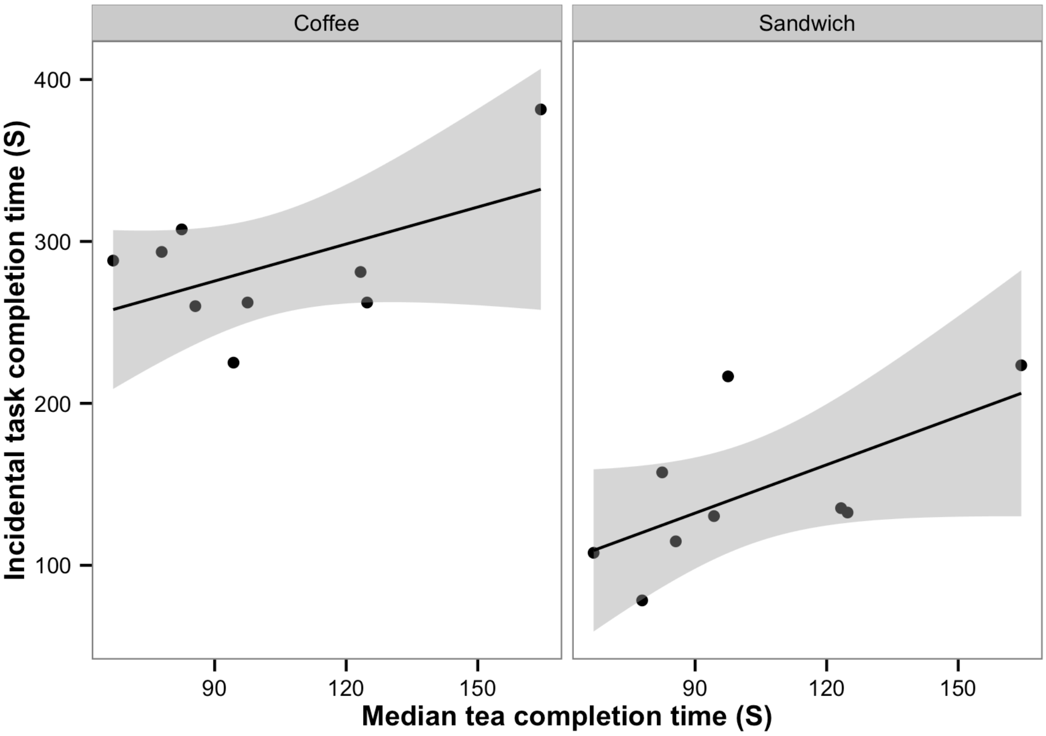


Figure 7.6 Median tea completion time correlated with incidental task completion time. This figure reveals a significant positive correlation, with participants who took longer to make tea also taking longer to make coffee or sandwiches.

Another way to explore this data is to look at the proportion of time spent looking at sandwich and coffee items during tea making. Figure 7.7 shows that when we look at the proportion of time spent looking at these items (compared to all other item types), that there is a significant negative relationship with task time being shorter with a higher proportion of looks (F(1,6) = 109.81, *p* < .001), and a significant interaction between proportion of looks and task type (F(1,6) = 7.193, *p* = .036).

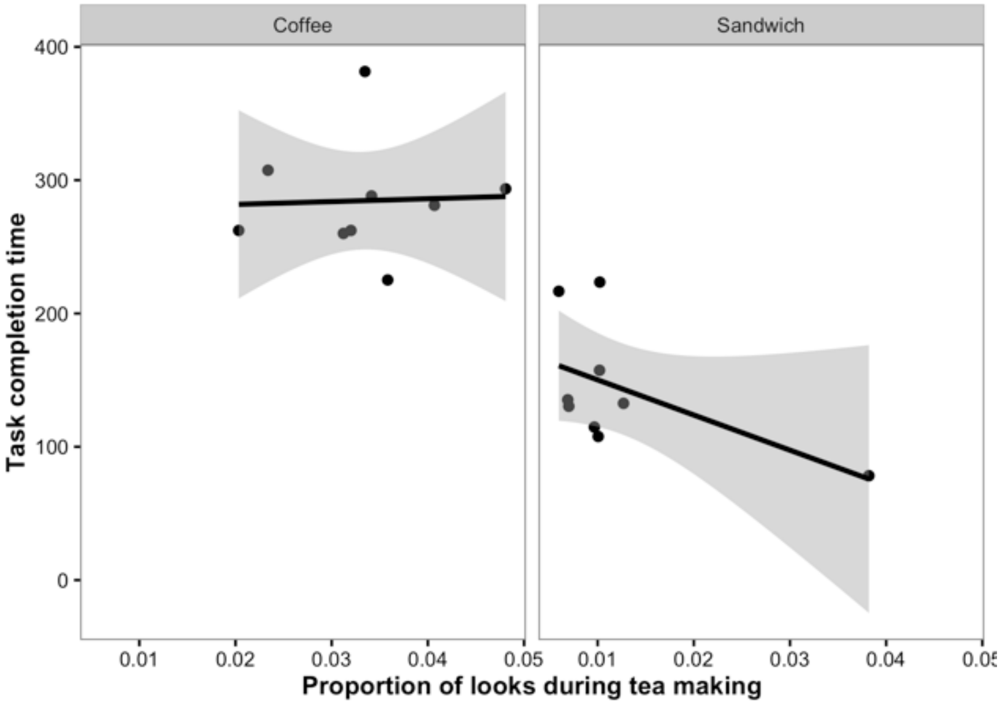


Figure 7.7 Proportion of looks during tea making that were allocated to objects of these task categories plotted against task completion time for coffee (left panel) and sandwich (right panel).

7.3.2.2 Exploration time and incidental encoding

We examined whether the proportion of the trial that the participant spent in the 'exploration' phase (i.e., not looking at task-relevant objects, not touching objects, and not in the eye-hand latency phase) was related to looks at objects during tea making using an ANOVA. There was no effect of task upon the amount of time spent in the exploration phase (*F* < 1), no effect of the number of looks during tea-making on the proportion of exploration time (*F* < 1) and no interaction (*F* < 1). When using the duration of looks to coffee and sandwich objects during tea-making, there was also no effect on the proportion of time spent exploring the environment (*F* < 1), with no effect of task (*F* < 1) or interaction (*F* < 1).

7.3.2.3 Look-ahead fixations and incidental encoding

There was a significant difference between the proportion of time participants spent looking ahead at objects in the coffee and sandwich tasks (F (1,7) = 8.196, p = .02), with participants looking ahead to coffee items for a higher proportion of the task. However, similarly to exploration time there was no effect of the number ($F < 1$) or duration ($F < 1$) of looks to objects for these tasks during the 10 days of tea making on the proportion of the current task spent looking ahead to objects that would be used in the near future.

7.3.2.4 Eye hand latencies and incidental encoding

As some coffee and sandwich objects were looked at more than others during tea making, we used LMMs to examine the relationship between eye hand latency of objects and the frequency, and duration of looks at these specific objects that occurred on the learning phase. The models included the task as a fixed effect, and participant and object as random effects. The maximal structures that converged for both models involved removing the term for correlations between intercepts and slopes. There was no relationship between frequency (β = -4.166, SE = 29.881, t = -0.139, p = 0.3) or duration (β =-0.4, SE = 0.924, t = -0.433, p = 0.84) of looks to objects during the previous 10 days and eye hand latencies when making coffee or a sandwich. In other words, it did not matter how much people looked at objects in the previous 10 days when they were not task relevant; this did not affect their eye hand latencies when they used the objects for the coffee or sandwich task.

7.4 Discussion

In the present study we were interested in whether exposure to objects which were semantically appropriate for the environment but task irrelevant were incidentally encoded and the subsequent effect on both general and visual behaviour, when a task switch meant that the previously irrelevant objects then became target objects required for an active task. The results from the previous chapter indicated that reliance on memory increased as participants became more familiar with an environment and its objects. During the ten tea-making days, the environment contained other objects, all of which were relevant in a kitchen but not relevant to the task in hand the present study then asked participants after this acquisition phase to complete tasks using these objects which had previously been present but irrelevant in order to examine whether or not task irrelevant objects were encoded incidentally. However, since we do not have the baseline data for visual exploration or looks to task irrelevant objects for making sandwiches and coffee without any incidental exposure, it is not possible to determine from these results whether or not behaviour changed as a result of incidental exposure to objects. It is possible that incidental encoding did not occur at all, or even if it did, exerted no impact on visual behaviour. Although it is not feasible to determine whether there was any facilitation of task completion due to incidental exposure, we can explore in a general sense whether the behaviour in the incidental tasks looks similar to that displayed for another task in a novel and familiar environment.

This question has been of interest in the literature previously and has set opposing views. It is possible that any effect of familiarity may be about representing the entire environment and content regardless of task relevance

(Hollingworth etc.,), or it may be that only task relevant information is represented (Hayhoe, 2003). The present study examined this issue, without using an explicit memory test. Since the purpose of the present study was to examine implicit encoding during an active task in a natural environment, we designed the experiment to examine memory in a more natural implicit manner thereby avoiding any issues of probing with for example alternative-forced-choice questions. To achieve this we used the same 10 participants who had made tea for 10 days in the same environment and on day 11 and 12 instructed them to complete a different task. The participants were naïve to the task switch and the kitchen was set up to resemble any other normal kitchen with semantically relevant distractor objects present, subsequently at the end of the 12 days none reported having suspected that there would be any task switches occurring.

During the acquisition of familiarity undoubtedly some information about the environment and the objects is being encoded but we cannot tell from the previous chapter what these representations include. It may be that the visual system represents only the information currently being attended to (O'Regan et al., 1999; O'Regan, 1992; Simons & Levin, 1997; Wolfe, 1999) or objects that are necessary for the task (Hayhoe, 2000) or that the representation is more inclusive and incorporates other objects and features of the environment that are encountered incidentally during viewing (Henderson & Hollingworth, 2003a, 2003b; Hollingworth et al., 2001; Williams et al., 2005). We found that the proportion of time spent visually exploring the room (which includes time spent visually searching for objects) was significantly different from day one of the tea making task but not from the proportion of time spent on day 10. People spent less time visually exploring the room than they had when it was novel, despite having to complete a new task with objects that had been visually present but never before interacted with. In terms of visual exploration, participants behaved more like they did when they were familiar with the environment. We also found that the proportion of looks to task irrelevant objects compared with looks to task relevant objects on task switch days were far less than on the initial day 1 of tea making, and in fact were similar to the proportion made during day 10 of tea making. These two broad measures suggest that some incidental encoding did occur during the 10 days of tea making, however whether this was encoding of the objects or the environment cannot be inferred from the measures above.

We can consider four possible explanations for what was being encoded incidentally during the acquisition phase, which may have contributed to the behaviour displayed on the two task switch days.

7.4.1 Could gist be all that is being incidentally encoded?

One of the most basic and rapid representations, we make about a scene refers to its overall meaning or nature, referred to as gist. In less than 100 ms (Potter & Levy, 1969) we extract this information independently of any explicit knowledge of the scenes details, such as the content or layout. It has been suggested that gist facilitates recognition of objects and enables the locating of informative regions in scenes (Brockmole et al., 2006; Brockmole & Henderson, 2006; Wu, Wick, & Pomplun, 2014) thus perhaps in the present study all that is being represented across the days is the gist of the scene and the reductions in visual search time and looks to task irrelevant objects are as a result of participants being able to use gist to fixate areas in the scene which semantically would be most likely to contain target objects. Draschkow, Wolfe, & Võ (2014) demonstrated that scene semantics aided search for objects in naturalistic scenes. The semantic guidance they refer to is a knowledge base of ‘scene priors’ which they argue, can be considered a form of memory and can guide search and support memory formation. In the present study if all that was being represented from the environment was the gist we would not expect to see a difference between day 1 of tea making and the two task swap days, since the gist was the same (i.e. a kitchen) across all of the days and tasks, however we did find that proportion of time spent visually exploring the environment and looks of task irrelevant objects (both features of search behaviour) were different to day 1 and in fact looked similar to the behaviour displayed when participants were familiar with the environment and objects. Gist may have contributed to the present findings, but cannot account for the reduction in these behaviours across days; therefore we must also consider another possibility of what information was incidentally encoded.

7.4.2 Representation of space

The other type of information that may be retained across task days is spatial representation in a general sense, Hochberg (1968) suggested that spatial layout information refers to the overall arrangement of scene items and features rather than the semantics and properties of objects. It may be that during the acquisition phase individuals are incidentally encoding information about the space in general and this spatial representation of the environment facilitates search, thus producing the lower proportion of time spent visually exploring and looking at task irrelevant objects. Kit et al. (2014) examined whether incidental fixations contributed to future searches for objects by presenting participants with a visual search task in an immersive virtual-reality environment on three successive days. They found that participants rapidly learned the locations of objects in the environment across time and argued that spatial memory was used to guide search, they also found no relation between the first search time and the number of incidental fixations. Kit et al argue that incidental fixations are not used to encode information rather general spatial information is represented and utilised to benefit subsequent search.

Further to the idea that participants were merely representing the space in a general manner relating more to the layout of the environment rather than the semantics of the rooms contents (Hochberg, 1968), it is possible that experience with the environment strengthened a contextual cuing effect (Chun & Jiang, 1998, 1999; Jiang & Wagner, 2004; Olson & Chun, 2002), Perhaps individuals were implicitly cued to target locations by learning the context of the scene (Brockmole et al., 2006; Brockmole & Henderson, 2006; Hidalgo-Sotelo, Oliva, & Torralba, 2005; Oliva et al., 2004) so for example learning that all things cuplike were stored in one cupboard may have aided search for the mug used only for coffee making, thus facilitating search and reducing the incidence of looks to task irrelevant objects. Instead of a process of illumination search strategy where each object in the room would have to be fixated and encoded serially, participants were able to use their knowledge of the environment to facilitate search thus reducing the visual exploration time and looks to task irrelevant objects.

7.4.3 Representation of objects

Whilst it may be that across the 10 days all that is being incidentally encoded is information about the gist of the scene or the spatial layout of the environment, it is also possible that we are encoding something about the objects present in the environment. It has previously been demonstrated that action influences perception (Calvo-Merino, Grèzes, Glaser, Passingham, & Haggard, 2006; Tatler et al., 2013; Wohlschläger, 2000) and memory for objects (Kirtley & Tatler, 2015; Tatler et al., 2013) even for prioritising specific properties of objects after performing an active task (Tatler et al., 2013), the present study investigates the issue from the opposite angle by exploring the effect of having looked and potentially incidentally encoded information about objects on the visual motor coordination during action. By examining the incidental fixations to objects that were task irrelevant during the initial 10-day tea-making task, we were able to explore potential effects on subsequent interaction with the objects when they became task relevant on days 11 and 12. Based on the findings of Chapter 6, we know that being familiar with an object changes our subsequent interactions with that object, however we do not know if the same holds true for objects that were merely fixated rather than interacted with. If we incidentally encode information about objects to the same level as in Chapter 6 we would have expected to see shorter eye-hand latencies much like those demonstrated on day 10 of tea making, however, we found no relationship between the number or length of fixations and eye-hand latencies when the coffee or sandwich objects were used on day 11 or 12. It may be however that the baseline eye-hand latencies of coffee and sandwich making here would be very different to those displayed in the tea making as the kitchen became more familiar thus we cannot conclude that these latencies reflect incidental encoding. Tatler & Tatler (2013) and Kirtley and Tatler, (2015) demonstrated that we prioritise encoding of specific information, particularly noting that shape was a priority when a person had to perform an associate action with the object. Properties like shape and orientation may be important for eye hand latencies since presumably one of the events during the latency between the eye and hand arriving is that the motor reach and grasp are being planned, thus the shape and orientation or position of objects would be prioritised for processing and perhaps even for encoding. If this was the case then we would have perhaps expected to see that some of this processing would have occurred during the fixations made during the fixations made in the initial 10 days of tea making, thus producing shorter eye-hand latencies, however as demonstrated by Kirtley and Tatler, (2015) the crucial element missing here was that during tea making the participants had no reason to suspect that they would subsequently have to interact with the task irrelevant objects (sandwich and coffee objects) this would presumably not have made an effort to encode details of the objects which could have perhaps impacted the visuomotor elements of the actions, like for example eye-hand latencies.

7.4.4 Evidence for no visual encoding

It is also worth considering that participants had not incidentally encoded anything about either the gist, the spatial layout or the environment or the details of the objects and in fact had only encoded the objects that were task relevant for the tea-making task. If this had been the case, the similarities between day 1 and the task swap days 11 and 12 for the proportion of time spend visually exploring the room and the proportion of looks to task irrelevant objects could be explained simply by participants having reduced their set size of object in the room that require fixating upon the task switch. By using their representation for the tea objects and their locations in the environment participants could have avoided looking at these which by default would have meant that their visual exploration time and looks to task irrelevant objects would have been reduced. If this was the sole cause of the similarities between day 1 and day 11 & 12, then we would have expected to see very few looks back to tea objects when completing the new tasks, however our results demonstrated that around fifteen percent of task irrelevant looks were to tea making items even although these objects were now task irrelevant. It is interesting to note that the percentage of looks to tea making objects was lower than the proportion of looks to task irrelevant looks on day 10 of tea making, consistent with the idea that some of the savings may be due to having encoded the tea making objects, thus allowing them to be avoided when they are no longer relevant. Although we can conclude that it is unlikely that simply cutting out looks to tea objects was responsible for the low visual exploration time and looks to task irrelevant objects, it may have reduced the proportion of time spent.

Overall our results indicate that there may be some level of incidental encoding occurring during familiarity acquisition, but that this may be restricted to representing the spatial layout of the room or perhaps even the semantic relationships of the particular environment, however without baseline data for both of the incidental tasks it is not possible to determine the extent of the impact incidentally encoded information has on subsequent object interactions. The data suggests that people behaved as though they were familiar with the environment in terms of locating objects but their behaviour did not indicate any changes in terms of actual interactions with objects as a result of having fixated or incidentally encoded them. This supports literature that suggests that initial task modulates what we encode and influences future visual behaviour (Kirtley & Tatler, 2015; Tatler & Tatler, 2013; Tatler et al., 2013; Võ & Wolfe, 2012). Essentially this study once again highlights the importance of environment and suggests that incidental encoding of an environment can aid subsequent visual behaviour, such as search, even if we are required to complete a new task which although familiar never completed before in the environment. Future work examining this issue is required to isolate the exact impact incidental encoding has on subsequent behaviour, baseline data for the task completion with no incidental exposure to the objects or environment would be required for any strong conclusions to be made.

Chapter 8 General Discussion

8.1 Summary of findings

The aim of Chapter 3 was to explore three crucial differences in visual behaviour during the completion of active tasks previously identified in the comparison of two separate similar studies by Land and Hayhoe (2001). The results from Chapter 3 demonstrated an effect of task type rather than environment, as had previously been speculated, with shorter eye-hand latencies when people were making sandwiches than tea, more unguided putdowns when making sandwiches and more looks to task irrelevant objects when making tea. Since the two tasks were similar both in terms of the semantics category (of domestic task) and in terms of the types of actions and manipulations required, it was proposed that the differences identified were most likely due to the objects used for each task. The results indicated that the eye-hand latencies varied between objects but that there was no variation related to movement around the environment.

Chapter 4 explored further the notion that the object properties impact visual behaviour, specifically visual guidance during the putdown of object was examined. The results revealed that the properties of the object, in terms of its style, material, fullness and information about the level or clutter of the area the object was being set down in impacted the likelihood of the putdown being visually guided or not. We found that in the first putdown of each glass, when the glasses were all empty the properties that seemed to matter were the style of the glass, whereas after having been filled, the level of liquid and the material also determined the level of visual guidance during putdown. As the tray began to fill with glasses, the clutter of the tray mattered in terms of deciding whether to visually guide the putdown, thus visual guidance was flexibly directed depending on the temporary state of the glass and its properties at that moment. Furthermore, it was identified that the term, guided putdown encompasses visual behaviour styles that could be further categorised into those that provide continuous visual guidance and those that still constitute visual guidance but not in a continuous manner. This was considered in the analysis and it was found that the same factors that influenced the likelihood of an object being put down with some visual guidance also contributed to the level of guidance provided. Since Chapter 4 identified no change across time in the impact of these factors, the results imply that participants are using some prior knowledge of the properties of objects and the way they would likely behave during putting down to base the decision of whether to support the action with visual or guidance or not.

Being familiar with the object types used in Chapter 4 afforded participants the ability to use visual guidance during the set down of an object based on the properties of the object. Chapter 5 furthered this line of enquiry by investigating the effect of being familiar or unfamiliar with an environment and objects on visual behaviour. The results revealed that visual behaviour during the completion of a familiar active task was different depending on whether the participant was familiar with the environment and the objects or unfamiliar. The results clearly demonstrated that whether or not an individual; was familiar with the environment and objects effected general behaviour, in terms of task completion time and visual behaviour, with regards to shorter eye-hand latencies for familiar kitchens, a larger proportion of time spent visually exploring the room in novel kitchens, making look-ahead fixations in familiar kitchens and in novel kitchens making more looks to task irrelevant objects.

The results from Chapter 5 indicated a clear effect of whether an environment was familiar or novel on visual behaviour however did not reveal anything about the development of familiarity. The purpose of Chapter 6 was to examine the process of familiarity acquisition and the changes in visual behaviour. We found that on day 10 of making tea in a kitchen, participants exhibited visual behaviours similar to those in the familiar condition of Chapter 5. Across the ten days participant became faster at the task, spent less time visually exploring the environment, looked less at task irrelevant objects and performed more look-ahead fixations. Furthermore there was a relationship between familiarity acquisition and the latency between the eyes and hand arriving on an object, with shorter eye-hand latencies as familiarity was acquired. We can conclude from Chapter 6 that during the acquisition of familiarity participants were encoding task relevant information which was then utilised to aid visual behaviour on subsequent days, however we cannot infer from that study what if any task irrelevant objects were incidentally encoded as a result of having been present – and possibly even looked at - during the completion of the task.

The aim of Chapter 7 was to understand what kind of information is being built up about an environment as we gain familiarity as a result of repeatedly completing an active task in the same environment. To do this we examined the effect of previous exposure to the environment and objects to see whether task irrelevant objects were incidentally encoded. By switching task after the initial 10 days of tea-making, we were able to examine whether participants had encoded information about the objects required for two other tasks (sandwich making and coffee making), both by the general exposure gained from experience with the environment and by making fixations to the task switch objects. We found that there was an effect on general visual behaviour: participants spent more time performing look-ahead fixations and less time visually exploring the room for both new tasks than they had for making tea on the first day of the tea making study in Chapter 6. However we did not find the shortened eye-hand latencies that seem to be consistent with having familiarity, therefore we concluded that either nothing was incidentally encoded and any changes in behaviour were due to the effects of becoming familiar with the tea making objects or that participants had incidentally encoded enough information about the environment that they required less visual exploration time and more time planning ahead. If the latter is correct, since the type of information encoded was not sufficient to change the temporal nature of the microstructure of the ORA, it may simply have been information about the spatial layout of objects that was encoded but that the information did not change the way participants interacted with the objects.

8.2 Comparison of findings

Within all of the studies presented in this thesis, certain measures of visual behaviour were taken in several of the chapters and therefore provide a means to make some comparisons between chapters. This is worthwhile for two reasons. First it allows us to ensure that there is some level of consistency in the findings across studies, and second, since the topics across all experiments interlink and all answer elements of the same overarching theme, being able to collate findings to identify common trends in the data means we can attempt to address this theme and identify what we still need to know in order to examine the effect of familiarity on visual behaviour during natural active tasks.

8.2.1 Eye-hand latencies

In terms of the sub-elements of active task completion, comparing results reveals that eye-hand latencies were affected by task (or more specifically the objects used in tasks), the level of familiarity with an environment and the acquisition of familiarity. They were mostly unaffected by movement and the microstructure of ORAs was unaffected by prior visual exposure without interaction when incidentally fixated. We found that in Chapter 3, the eye-hand latency for making tea was approximately 720 ms, whereas for making sandwiches it was 570 ms. Results from Chapter 5 revealed that in familiar kitchens eye-hand latencies for tea making were around 640 ms. This is much shorter than the 750 ms eye-hand latencies when participants made tea in a novel kitchen. Eye-hand latencies for Chapter 6 revealed that the average eye-hand latency for day 1 was 620 ms whereas for day 10 was 586 ms. The latency with which the eye leads the hand in Chapter 7 was found to be 722 ms for coffee making and 660 ms for sandwich making. We can see that in novel kitchens (Chapters 3 & 5) the eye-hand latency for making tea was very similar, and the latencies for Chapter 6 are consistent with this. In particular we can see that at the start of the ten days eye-hand latencies resembled those found in Chapters 3 & 5 for novel kitchens then reduced to look far more similar to the familiar kitchen latencies in Chapter 5.

8.2.2 Guidance during putdowns

Visual guidance during the putting down of an object was affected by the task being completed and by the types of objects being set down. We found that in Chapter 3 there were more guided putdowns when people were making tea (48%) compared to making sandwiches (65%). In Chapter 5 12% of putdowns in the novel kitchen were unguided, whereas 10% were guided when the kitchen was familiar. In Chapter 6 on day 1 of tea making the percentage of unguided putdowns were 30%, this was reduced by day 10 to 20%. Finally, in Chapter 7, 24% of putdowns were unguided when the task was coffee making, compared to 29% for making sandwiches. There is some inconsistency with these results, the nature of which will be unpacked in this discussion.

8.2.3 Looks to task irrelevant objects

Looking at task irrelevant objects seems to have been affected in Chapter 3 by both the task and the layout conditions of the environment, with 11% of gaze events being directed to irrelevant objects when making tea compared to only 3% when making a sandwich and 11% when objects were laid out apart but only 3% when objects were together. Chapter 5 revealed that in the novel kitchen 39% of looks were to irrelevant objects whereas, in the familiar kitchen irrelevant objects only accounted for 17% of the looks. Chapter 6 revealed a similar trend if we consider day 1 ads novel, with 58% of looks to task irrelevant objects compared to day 10 when that reduced to 26%. In Chapter 7 we found that even although there had been a task switch the proportion of looks to task irrelevant objects after having spent 10 days acquiring familiarity looked similar to the low proportions in familiar kitchens, with 18% for tea making and 32% for sandwich making.

The comparisons here indicate that not only are the findings consistent across chapters but that there are effects of familiarity level with an environment and objects on visual behaviour during the completion of an active task. Shorter eye-hand latencies and fewer looks to task irrelevant objects seem to be features of having acquired familiarity with an environment and objects, however if actual experience of certain objects is not acquired during the acquisition phase there is no subsequent affect on sub elements of ORAs, such as eye-hand latencies. The way in which these findings fit with previous literature and contribute to our knowledge of the topics will be the aim of the rest of this discussion.

8.3 Searching for objects

8.3.1 What did we know?

For search tasks using natural scenes we know that search is guided not only by features of the target objects (such as size and colour) but also by scene specific factors, such as sematic and structural knowledge of the scene By using our general semantic guidance accumulated and collated from many experiences, we can develop a schema for typical kitchens and the rules that apply in kitchens, for example that kettles are usually placed on worktops, and are thus able to reduce the set size of potential targets and areas on a scene or environment to search. (Bar, 2004; Biederman, Glass, & Stacy, 1973; Eckstein, Drescher, & Shimozaki, 2006; Torralba, Oliva, Castelhano, & Henderson, 2006; Võ & Henderson, 2009; Wolfe, Alvarez, Rosenholtz, Kuzmova, & Sherman, 2011)

We can also use ‘episodic guidance’, which refers to specific knowledge about a particular scene to guide search to areas of a scene or environment where we know the object is located (See Wolfe et al., 2011 for review). However, episodic guidance can only be acquired if a viewer has had the opportunity to accumulate sufficient experience with the specific scene (Võ & Wolfe, 2011). Several authors have examined search after repeated exposure to a scene and have found that there is an effect of subsequent search behaviour ((Hollingworth, 2006, 2009; Hollingworth and Henderson, 2002). Brockmole and Henderson (2006) repeatedly showed photographs of scenes that contained an embed target (a letter T or L) that were consistently but arbitrarily placed. The images were then flipped to a mirror image. Observers initially looked at the original (expected) location for the target object but then quickly moved to look at the new location, although they were slower than when the object was in the initial place, search time was still faster than if viewers were completing search of a new search. Similarly Võ and Wolfe (2012) showed that if a subject searched for an object in a scene, the subsequent search for the same object was speeded dramatically despite many intervening searches. These studies demonstrate that experience with a scene can impact our visual behaviour, even to the extent that we rely on our expectations to make anticipatory eye movements to where we think an object will be, rather than use the visual information available in the very first instance (Brockmole & Henderson, 2006).

In the course of searching for targets, particularly where we have no episodic knowledge of the scene, there are a number of looks to task irrelevant objects. In many real life instances, search is an element of a task embedded in another task, for example when making tea in a novel environment, searching for the tea-making objects is part of the task but not the main task goal. Research examining search related visual behaviour during the completion of an active task suggests that individuals make relatively few looks to task irrelevant objects. Land, Mennie and Rusted (1999) report that only 5% of looks are to task irrelevant objects, whilst Hayhoe (2003) noted that prior to the commencement of the task 48% of looks were to task irrelevant objects but that as soon as the task began this reduced to 16%. This drop in the number of looks to task irrelevant objects is an interesting finding, and is a clear demonstration of the importance of task on our visual behaviour.

8.3.2 What do we know now?

The data examining visual exploration collectively indicate that behaviour changes as a result of familiarity with an environment (Chapters 5, 6 & 7). The fact that all participants were able to complete the tasks without difficulty and that no-one looked at areas of the kitchen that were unlikely to contain the necessary task items (for example, no participants looked to the ceiling or floor to locate the kettle) suggested that people are able to utilise prior knowledge of the semantic structure of the scene to aid visual search in a novel environment. Task completion time was considerably longer in the novel environment than in a familiar environment (Chapter 5) or in the same environment once familiar (Chapter 6). People spend a fairly small amount of time visually exploring compared to when they are unfamiliar with the environment. Although we cannot say for sure, since the spatial extent of searches was not measured, the reduction in time could be an indication that participants were able to draw on episodic guidance to aid search in familiar environments. Since extensive visual searches to locate objects were not required in familiar kitchens, it may be that people performed less visual exploration of the environment in general. Proportionally less time spent visually exploring the environment for familiar environments also emerged as a trend across days during the acquisition of familiarity in Chapter 6. Parallels between the results from Chapter 5 and 6 indicated that visual exploration (in terms of the proportion of time) during the first day of the tea-making task, looked similar to the novel kitchens in Chapter 5, and by day ten looked significantly different, instead resembling the exploration displayed for novel kitchens in Chapter 5. The fact that exploration time decreases implies that some information is being retained and allows for a decrease in the amount of time needed to visually explore the environment, this suggests that memory for the environment and objects is being built up and used to facilitate search. During the acquisition of familiarity with an environment and its objects, the information encoded may be represented in memory, which then allows us to utilise episodic memory to locate and retrieve target objects. The reduction in proportion of time spent searching across task days may represent this increasing reliance on episodic guidance.

To reconcile this finding with the literature suggesting that viewers prefer to make eye movements to search anew, rather than rely on memory (Oliva et al., 2004), it is beneficial to consider the search element of the task imposed by the environment in this thesis as more akin to foraging tasks. The literature on foraging (Gilchrist et al., 2001; Smith et al., 2008) suggests that the high costs in terms of effort to physically search for visually unavailable information mean that people rely on memory more, particularly in order to avoid fruitless revisits. All of the kitchens used in Chapters 5 & 6 and 7 had information (target objects) which were hidden from sight, in cupboards and drawers or even behind other objects, essentially requiring foraging in order to be located, thus utilisation of the episodic guidance being built up across the 10 days, or from previous experience in Chapter 5 would have in turn produced the reduction in proportion of time spent visually exploring the environment. This would have been particularly true in situations where the visual information was not available without physical search. As discussed in Chapter 5, the relative reliance on vision vs. memory depends on the quality of available visual information (Brouwer & Knill, 2007), if the information available is not present then having better memory for objects would influence search since the weighting in this case would favour memory over vision.

For all of the experiments where looks to relevant and irrelevant objects were measured, we found that during the completion of an active task in a novel environment (Chapters 5 & 6) around 49% of task time was spent looking at irrelevant objects, whereas, this was only 23% for familiar environments (Chapters 5, 6 & 7). This finding implies that looks to task irrelevant objects reduces if the environment and objects are familiar and implies that they have been encoded in memory which in turn suggests that in familiar environments memory is more heavily weighted as the strategy to be used for locating objects. Perhaps in becoming familiar with an environment one does not only encode where to look (locations of target objects) but also where not to look (locations of task irrelevant objects), using episodic guidance to locate objects people spend less time visually exploring and therefore make less looks to task irrelevant objects. Again this is particularly true due to the characteristics of the environment used for these experiments, as already discussed for Chapters 5, 6 & 7 many of the objects necessary for the task were not within plain sight, thus here familiarity affords the individual the ability to weight memory to facilitate search, rather than having to physically search the environment for each target object.

The layout of an environment impacts visual search, support for this comes from differences found between Chapter 3 compared to Chapters 5, 6 & 7. In Chapter 3 looks to irrelevant objects were only 11% for tea and 3% for sandwich making, clearly considerably lower than for the novel kitchens in Chapters 5 & 6. The crucial difference here was that in Chapter 3 all of the objects were within plain view, so the looks to task irrelevant objects were possibly low as a consequence of not having to search, whereas in the novel kitchens in Chapters 5 & 6 locating several of the objects required physical search. In Chapter 3, looks to task irrelevant objects in the novel environment were minimised due to the layout of the environment and the fact that all objects were very easily locatable. Whereas, in the novel kitchens in Chapters 5 & 6, the objects were hidden from view requiring physical search to locate, thus increasing the incidences of looks to task irrelevant objects. In familiar kitchens (Chapters 5, 6 & 7) the looks to task irrelevant objects reduced significantly (although were still slightly higher than for Chapter 3) not because the environment change, rather the change was in the better memory of the environment and objects and the increased ability to rely on this memory rather than the visual information, thus reducing the number of looks to task irrelevant objects.

Ultimately then we can see from out results that search is not only sensitive to the layout of the environment but also our level of familiarity with it. More specifically, if the environment does not have all of the information in plain sight, having familiarity (better representations of the information) facilitates search by allowing the locating of objects to rely more heavily on memory. In situations where we are unfamiliar with an environment the amount of search and indeed looks to task irrelevant objects we conduct will depend on whether the objects are in plain view or require visual search.

8.3.3 What do we still need to know?

Due to the self directed nature of the sub goal completion for all of the experiments presented in this thesis, only an overall measure of 'visual exploration' rather than search can be made. Typically in visual search tasks, targets are searched for serially, that is to say participants are instructed to search for an object or a change, thus fixations made or time taken to locate the target is a reliable measure of search. In tasks where search is a sub-element of an active task in which some or all of the sub-goals could be ordered in different ways, it cannot be assumed that the next object acted on was the intended search target. Due to the self scripted nature of the task, we cannot infer what the person intended to search for next: for example, it is entirely plausible that they may have intended to search for the teapot but during that search, fixated on another potential target object and simply adjusted the order of sub-goals to fit so that this fixated object was used next. In instances like this we would no longer be measuring just the amount of time it took to search for one object. Within that time frame there would have been at least one target search, fixations that happen to land on another target object, discrimination to ensure the accidentally fixated object was a target, and potentially adjustment of sub-goal script. There is no way to quantify the number of times this occurs during the completion of a task, so as a means to address this issue the present study used the term visual exploration to include all looks to objects that were not attached to an ORA of look-ahead fixation. The proportion of time spent visually exploring the environment was affected by the level of familiarity a person had with the environment and objects which suggests that search times may be faster in familiar environments. Thus an interesting follow-up experiment would be to isolate time spent searching during an active task. This could be achieved by manipulating the flexibility of task scripts, to remove the self directed element so that there would be no opportunity to change sub-goal order based on accidental fixations to task relevant objects.

8.4 Extracting and retaining information

8.4.1 What did we know?

For the completion of an active task, typically the information we need changes across time, as each task sub-goal is completed and another is begun. The strategy used to ensure fixations are guided to informative areas and objects may depend on the context of the task. If the effort involved in making eye movements to gather information was relatively low, for example if all the targets are on a screen or a surface in front of the person, one strategy would be to rely heavily on the visual information available in an online manner, taking only the information required as and when needed, (Ballard, Hayhoe, Li, & Whitehead, 1992; Ballard, Hayhoe, & Pelz, 1995). Another strategy would be to utilize prior knowledge gained through experience to fixate areas of interest where the required information is likely to occur. For example medical experts have been shown to use their experience of inspecting images and locating abnormalities to quickly and accurately fixate and inspect locations where the required information is likely to be (Krupinski et al., 2006; Stainer et al., 2015). The strategy of utilising prior knowledge may be of use when there is a time or accuracy pressure to have fixations extract information from a specific spot, thus reducing the set size or area of search by utilising prior knowledge would facilitate search. Whilst these strategies may be used in instances where objects behave fairly predictably, there are also instances where the activity is more dynamic and unpredictable. In these instances, it is likely that both elements of online information extraction coupled with prior knowledge of where the information is likely to be are used in order to guide the eyes to the vantage point. Expert sports players and drivers have been shown to be able to do this extremely quickly. Expertise in sport particularly, appears to aid accurate online evaluation of the visual information at a speed which in turn allows fixations to be proactive, fixating an area which is about to become informative (Land & McLeod, 2000; Land, 1992; Land & Tatler, 2001; McKinney, Chajka, & Hayhoe, 2008; Underwood, Chapman, Bowden, & Crundall, 2002; Underwood, Chapman, Brocklehurst, Underwood, & Crundall, 2003a, 2003b; Vaeyens, Lenoir, Williams, Mazyn, & Philippaerts, 2007). Typically, these types of studies demonstrating the strategies of experts are unable to reveal anything about the process of becoming familiar with a task, the progression from an online, as-needed strategy of information extraction approach to one based more on utilising prior knowledge. The process of learning and the transformation of visual behaviour was demonstrated by Sailer, Flanagan, and Johansson (2005) who found that during the learning of a novel task participants fixation strategies revealed a three-step process of learning where it was not until the third stage of skill refinement that participants were able to shift gaze ahead of the target.

It is not just information taken from task relevant objects that illustrates the extent to which we extract information and then use to build representations. Looks to task irrelevant objects can reveal important evidence about the amount of information that we incidentally encode. It may be that only the information necessary for the task is represented (Hayhoe, 2000) or that the representation is more inclusive and incorporates other objects and features of the environment that are encountered incidentally during viewing (Henderson & Hollingworth, 2003a, 2003b; Hollingworth & Henderson, 2006.; Hollingworth, Williams, & Henderson, 2001; Williams, Henderson, & Zacks, 2005). From the literature we also know that the level of what is encoded about objects may change depending on the task that was being completed (Tatler & Tatler 2013; Tatler et al., 2013; Kirtley & Tatler, 2015; Võ & Wolfe, 2012). What we don't know is when individuals are completing an active task in an environment, is information from incidental fixations retained and if so what would be the impact on visual behaviour when the task relevant objects then became relevant.

Regardless of whether the task is purely a visual search task or whether there is an element of search as a feature of another task, the amount of information incidentally encoded from looks to task irrelevant objects is of theoretical interest since it can illuminate what is being encoded from a scene or environment, and whether search is supported by increasing knowledge of task irrelevant as well as task relevant objects and locations. Exploring what information is retained from these incidental fixations has suggested that even for task-irrelevant objects simply looking at an object can indeed result in memory retention (Castelhano & Henderson, 2005; Hollingworth, 2006; Williams, Henderson, & Zacks, 2005). It may be that only the task relevant objects are encoded and represented (Hayhoe, 2000), or it may be that we are representing something about the spatial layout without knowledge of the content (Hochberg, 1968) or even the relationship between objects in the scene or environment (Rosman & Ramamoorthy, 2011). Alternatively we may be encoding the irrelevant objects themselves, their locations and even their properties and features (Golomb, Kupitz, & Thiemann, 2014) Considering the literature it is unlikely that it is none of these factors and that we truly retain nothing (O'Regan, 1992). We know then that visual search for scenes is affected by having had prior exposure, and we know that during the completion of an active task a small proportion of our visual exploration time is spent fixating task irrelevant objects but we cannot tell from the literature whether the level of familiarity one has with an environment and its objects changes visual behaviour in terms of visual exploration, looking at task irrelevant objects and incidental encoding during an active task in the same way as occurs for visual search of scenes.

In an active task the extraction of information is not just performed for the object currently being used, instead it is often the case that information is being extracted for objects that are to be used later in the sequence. These looks to objects, termed look-ahead fixations, have been found in several sequential tasks, including hand-washing (Pelz & Canosa, 2001), model building (Mennie et al., 2007), tea and sandwich making (Land & Hayhoe, 2001) and in a speed cup-stacking experiment (Foerster et al., 2011a). The consensus as to the utility of look-ahead fixations seems to be that they aid planning of a sequential task, Mennie et al. (2007) suggest that look-ahead fixations are used to plan the next movement, both in terms of planning the fixation and reach upon the initiation of a manipulation. We do not yet know if the frequency of look-ahead fixations changes for an environment that is familiar and what this could suggest for planning. For example the proportion of time spent making look-ahead fixations reduces in familiar environments, so perhaps this is evidence of building representations of the environment and objects. The information acquired during the acquisition of familiarity i.e., the representation, could then be utilised and planned from rather than having to use fixations to support planning during the actual task.

8.4.2 What do we know now?

Chapter 7 revealed that we do seem to encode something about the environment and objects contained in it even if we have only incidentally fixated them. After having made tea in a kitchen for ten days and then switching tasks on two subsequent days, (to coffee and then sandwich making, or vice versa), participants spent proportionately less time visually exploring the room (an indication of search time for target objects) and more time making look-ahead fixations. Thus the data looked more like the data someone completing a task in a familiar kitchen (Chapter 3, 4, 5 & 6) than a novel one. From this result we can conclude that something about the environment and objects was encoded during incidental fixations but that this may be limited to encoding of only the representations of locations of objects rather than any specific details of the objects. Further to this, the results from Chapter 7 indicated that there was no impact of having previously incidentally fixated irrelevant objects on the subsequent visual behaviour and manipulations. There was no relationship between having looked at the object (in terms of number of looks or time spent fixating) and the temporal aspects of the ORA. This result suggests that the information about an object incidentally fixated that may speed up elements of ORAs is not committed to memory and rather the information that is extracted and encoded is related to the structure of the scene.

The results from Chapters 3, and 5 indicate that the proportion of time spent making look-ahead fixations changes with the level of familiarity. More look-ahead fixations are made when a person is familiar with the environment, and as familiarity is acquired other behaviours decrease but the proportion of time spent making look-ahead fixations increases. We can establish that look-ahead fixations in this setting were unlikely to be only related to planning the deployment of the saccade to the next object since typically locomotion or at least a head movement may have been required to arrive at the next object, however it is plausible that as argued by Mennie et al. (2007) planning the saccade in terms of the direction to launch the fixation to, is planned during look-ahead fixations. It may be, that as suggested by Pelz and Canosa (2001), we make look-ahead fixations in order to aid the transitions between different elements of the task, thus look-ahead fixations may reflect something about the planning of the order of sub-goals to be completed. Perhaps the utility of look-ahead fixations is to ensure that the next required item is in location or to re-affirm that the next designated target object is a logical next move.

The data from Chapter 5 indicated that the order the flexibly scripted sub-goals were completed in was not random every day (for example people didn’t tend to switch the kettle on at a different point in the task each day), the task script got more similar across days. This suggests that people are finding a consistent way to make tea that is refined over days, not simply repeating what they did on day 1. The results from Chapters 5, 6 and 7 may all suggest that familiarity with an environment and objects offers some of the same advantages as familiarity with a task does, in that after a certain level of skill has been achieved (in this case learning the layout and ideal order of actions based on the environment) resources are freed up from visual exploration and instead can be deployed to plan ahead. The organisation of behaviour appears to be flexible in terms of how actions are ordered. We are able to achieve the same task even in an environment where no existing knowledge of the environment or objects exists and are able to refine the order of our actions as we acquire familiarity.

8.4.3 What do we still need to know?

The conclusions drawn above relating to the utility of look-ahead fixations as reflecting some level of task planning are speculative and were not a deliberate manipulation of this thesis. In order to categorically define the utility of look-ahead fixations we could examine the exact locations on objects of where look-ahead fixations initially land, for example to examine whether they facilitate planning of motor action to pick up the object or if they reflect planning about the manipulation to be performed with the object then based on the literature (Desanghere & Marotta, 2011; Flanagan, Bowman, & Johansson, 2006) the fixations would either be concentrated on the main body of the object for the former notion or on the area of functional activity for the latter idea. If we were to understand the utility of the look-ahead we would gain valuable insight into what is being extracted and represented from these fixations, and subsequently used for planning.

With regards to the order sub-goals are performed in self –directed tasks, we do not know if the changes in sub-goal completion reflect a persons preferred order or if it tells us something about the layout of an environment and the learning of that. It may be that as participants learn their way around a space and the locations of the objects they adjust their script in order to perform more efficiently in the space, or to avoid the effort involved in missing these opportunities and having to back-track to a target object close to a previously manipulated one. Presenting a task where the order of sub-goals to be completed and the opportunities and effort in different layouts and order types was manipulated would begin to address this issue. By investigating whether the way we perform a task is dependent on restrictions in the environment or based on self directed preferences, the flexibility of our capacity to update a task script based on changes in the environment and the subsequent effect on visual behaviour and performance would be revealed.

8.5 Object related actions

8.5.1 What did we know?

In Chapter 3 the importance of understanding the microstructure of the ORA if we are to understand complex behaviour was discussed. We know that practice with a task can speed up task performance, (Crossman, 1959; Epelboim et al., 1995, 1997; Foerster, Carbone, Koesling, & Schneider, 2011b; Herst, Epelboim, & Steinman, 2001; Newell & Rosenbloom, 1993; Schütz-Bosbach & Prinz,

2007) the decrease in completion time could be as a result of, not only the reductions in time spent searching and looking at task irrelevant objects, but also either the actions themselves becoming faster, the contraction of visuomotor co-ordination (for example reduced eye-hand latencies), or a combination of all of these. It is widely acknowledged that during an active task the eye leads the hand, typically by around a second (Land, Mennie & Rusted, 1999). However, variation in this lead-time of the eye to hand has been found, in sandwich making the average latency was around 100 ms whereas, for the high speed task of cup-stacking (Foerster, et al., 2011) the eye-hand latencies were found to be much more similar to that of tea making. One of the factors that changes the latency with which the eye leads the hand is learning. Sailer et al. (2005), Epelboim, Steinman & Kowler et al. (1995), Epelboim, Steinman & Kowler et al. (1997), Johanson & Flanagan (2003), and Foerster et al, (2011) have all demonstrated that as people become faster at the completion of a task, either with or without an externally imposed time pressure, eye-hand latencies get shorter. We can assume therefore that the latency between the eye and hand arriving on an object at the commencement of a manipulation is not fixed. If some level of processing is occurring during this latency in order to programme the appropriate reach, touch, or manipulation, it may be facilitated by having learned the task and associated actions. We do not know form previous work whether these elements of the ORA, i.e., the eye-hand latencies and the pace of task completion change even for a familiar task when the level of familiarity with the environment is different. Investigating these issues would allow the exploration of the factors that govern spatiotemporal coordination of vision and action within complex real world activities.

8.5.2 What do we know now?

In this thesis we have found that various aspects of the ORA microstructure change as we become familiar with an environment and objects. We find that even for familiar tasks the eye-hand latencies are affected by the task type (Chapter 3) and by the level of familiarity with the environment and objects (Chapters 5, 6 & 7), with shorter eye-hand latencies in familiar environments. Since neither the object types nor the manipulations to be completed with them were novel, the effect of familiarity for Chapters 5, 6 & 7 must have arisen from participants learning of the environment and the specific details of the objects. This raises questions about why and what might be going on during the eye-hand latency, if it was simply about getting the eyes as far ahead as possible, as might be predicted based on the results of Sailer et al. (2005) then we would expect the opposite direction for eye-hand latencies in the familiarity study. However the fact that they decrease suggests that perhaps some process within the eye-hand latency is becoming streamlined. This may be related to the amount of processing that occurs during the eye-hand latency suggesting that in familiar environments, less processing was required during the initiation of an ORA, or that the same amount of information was processed quicker, therefore the necessity of a longer latency between the eye and hand arriving in order to programme the motor response was reduced.

Further to there being a reduction in eye-hand latencies for familiar environments, other aspects of the ORA also got shorter. The latency between the eye arriving on the object and the change of state occurring was lower for familiar kitchens (in both Chapters 5 & 6). Similarly the latency between the arrival of the hand on an object and the change of state was lower when the kitchen was novel (this was significantly lower in Chapter 5 but not so in Chapter 6, although there was still a reduction of 138 ms when the participants made tea in their own kitchens). In Chapter 6 it was found that even the time from the start to the end of the COS changed, reducing across days suggesting that even for a task that is already considered familiar the physical element of object manipulations is sensitive to the level of familiarity with those specific objects.

It was previously proposed by Land et al. (1999) that that placing a familiar task in a novel environment would only effect the ease and speed that objects are located, and not the fluidity of the actions performed upon them, however the results from this thesis suggest that the microstructure of the ORA and the temporal nature of these elements of action are indeed sensitive to the familiarity of the environment and the objects in the environment. If these elements of the ORA change depending on the context they are performed in, regardless of familiarity with the task, this could imply that they are not fixed with regards to timing for even familiar tasks. This flexible temporal nature of latencies and element completion times of the ORA could have implications for all behavioural measures of active tasks. Tasks may be speeded up or slowed down depending on whether the environment and objects are familiar or not and the effect this has on changing the timings of the microstructural elements of the ORA. In a broader context, if the point at which the eyes fixate the object, the length of this fixation before the arrival of the hand, and the time taken to manipulate the object the object are significantly faster when we are familiar with an environment, then this may have implications for the propensity of conducting experiments in unfamiliar environments. Establishing the factors that can affect the microstructure of the ORA and the limits of temporal flexibility is important if we are to continue to apply findings from experimental setting to the way we behave in everyday life.

8.5.3 What do we still not know?

If the utility of the eyes leading the hand is related to planning the touch, reach and or manipulation (Batista, 1999), then an interesting question to follow up with would be, what performance advantage does advanced processing affords us during the completion of an active task?. It is difficult to present a natural task, but remove opportunity to lead with the eyes at the start of an ORA, however most of the related literature acknowledges that there are instances where even for these natural tasks the eyes and hand arrive simultaneously or even cases where the hand arrives before the eye (Land & Hayhoe, 2001). Further study of these instances, either naturally occurring cases or as a result of an experimental manipulation, and the subsequent impact on ORA performance would begin to unpack this issue. If we could design a task where there was no opportunity to have the eyes arrive before the hand on the object (thus removing the eye-hand latency) we could examine whether performance is impaired. If the fixation time is about planning the reach or touch, then removing the opportunity of early arrival of the eye might make the reach less accurate. Or it may be that the purpose of the eye arriving before the hand might be to do with the actual manipulation, Hayhoe, Shrivastava, Mruczek & Pelz (2003) noted that the locations of the fixations on the objects were different for different actions, for example, the middle of the jar for grasping, and the rim for putting on the lid, suggesting that fixations appear to play a specific role, depending on momentary task context. Thus removing the opportunity to have any extra processing performed during the eye-hand latency here would perhaps impair the actual manipulation. Given that it seems to be a fundamental building block of active vision, that the eye leads the hand during the completion of an active task, establishing what exactly is being processed during this time would reveal crucial information about the spatiotemporal allocation of gaze and the purpose of vision being active.

8.6 Visual guidance at the end of ORAs

8.6.1 What did we know?

We know a lot about the amount of visual guidance directed to an object both at the commencement of an ORA and during the manipulation of the object. We know that the eye tends to lead the hand at the beginning, allowing for some motor planning (Land et al., 1999; Hayhoe, 2000; Foerster, et al., 2011), and that during a manipulation one of the purposeful eye movements made is guiding the action, for example guiding the knife with jam on it to the bread. We even know that at the end of an ORA when setting down an object, we often guide the set down by fixating an empty location in the environment, for example an empty spot of worktop and fixate there until the object has made contact with the worktop, thus guiding the putdown of the cup. However, it has also been shown that there are instances where this guidance during the set down of an object does not occur (Land et al., 1999; Hayhoe, 2000; Land & Hayhoe, 2001). Little evaluation of why there may be variation in whether an object is guided during the put down or not has been conducted. In particular there has been no research conducted on the factors that mean some objects are visually guided during put downs but others are not. We can infer from work examining the factors that influence visual behaviour at the beginning of interactions with objects that object properties may have some bearing on our visual treatment of them after the functional manipulation is complete. Cole (2008) demonstrated that prior knowledge about the height combined with online evaluations of visual information to update priors were used to plan the lifting of objects and that even when there properties were subtly changed (so subtly that participants did not report noticing) the online assessment updated the information so seamlessly that the motor act of reaching, grasping and lifting was adjusted accordingly. Thus if the properties of objects can impact our behaviour at the beginning of actions it is conceivable that the same could be true for the putting down of an object after the action manipulation is over. One study that implies that there may be differences in visual behaviour towards the end of a manipulation was carried out by Sims, Jacobs, & Knill (2011) who used a block-sorting task presented in a virtual workspace environment, which required vision to be used both for information acquisition and on-line guidance of a motor act. The difficulty of the task was varied at the point of the task in hand and for the next action to be completed. Sims et al. (2011) argue that participants adaptively adjusted fixation allocations and durations based on the difficulty of both of these factors. One of the difficult elements of the task manipulated by Sims et al, was in terms of the bins the blocks had to be placed into. Their findings revealed that placing a block into a smaller bin required more visual guidance. Thus the properties of the objects being used in a task like this have a bearing on the level of visual guidance implemented when completing actions with them. Some of the central conclusions about vision during an active task have been primarily based on how we behave at the initiation of an interaction with an object and during the manipulation but if we are to understand the whole of vision for action then we must establish what happens at all stages of the actions, including the behaviours that emerge at the end of an action, particularly since in sequential tasks the end of one action directly affects the commencement of the next.

8.6.2 What do we know now?

All chapters in this thesis demonstrate that there are both instances of object putdowns that are visually guided and visually unguided. In Chapter 3, 36% of putdowns of objects were performed without visual guidance during tea making whereas during sandwich making 52% of objects were unguided during putdowns. Since the environment was consistent and there was no effect of moving around the environment, the crucial factor that differed between the two tasks was the objects and related manipulations. Further analysis revealed that there was also an effect of object type on the instances of objects being put down without visual guidance. Chapter 4 examined this issue further by deliberately manipulating the properties of objects to observe the effect on the level of visual guidance during the setting down process and found that the level of guidance delivered is flexible depending on the properties of the object and in fact the areas it is to be set down in. The same object can be the subject of different types of visual guidance depending on it’s temporary properties at the time of manipulation, so that when glasses were empty, height appeared to be the determining factor as to whether the putdown would be guided or not, with tall champagne flutes being the most likely to be guided and short tumblers being the least. When the same set of glasses were then filled with liquid to varying levels, the factor that influenced visual guidance during set down was fill level and material, with those made of glass rather than plastic and full glasses rather than empty or half-full ones being more likely to be visually guided. Again, the flexibility of this was demonstrated when the area of set down became more cluttered, in this case presumably to avoid collision when putting the glasses down, as the clutter of the tray increased so too did the likelihood that the set down would be guided. Chapter 4 also revealed that within visually guided putdowns there are at least two categories, those putdowns that are continuously visually guided and those that receive non-continuous guidance; the level of guidance was also sensitive to the same factors as mentioned above, with continuous guidance more likely to be directed at tall glasses, those full of liquid and at glasses being putdown on a cluttered surface.

Examining visual behaviour at the end of ORAs is an important endeavour not only because it is as much a part of the task as the other stages of ORAs but also because if we are interested in the temporal relationship between ORAs then one of the most important factors that can impact the subsequent actions in a sequence is the point at which the eyes are free to leave the current manipulation and move on the next. If all of the objects in a task require visual guidance during the set down or vice versa then the impact this would have on the pace of the task in general, the eye-hand latencies and perhaps even the time spent making look-ahead fixations would all likely be affected.

8.6.3 What do we still not know?

Chapter 4 in particular goes some way to teasing apart the factors that may influence the way we visually behave during the putting down of an object, however we cannot tell from our data what influence the same factors had on our general motor behaviour. If we use visual guidance to avoid risk, there may also be changes in motor behaviours which occur concurrently, for example visually guided put downs may be slower since presumably more care is being taken or alternatively it may be the opposite, perhaps when confidence of achieving a put down with no visual guidance is high this conviction would be apparent in the motor behaviour also, with faster put downs as a result. Data examining changes to behaviour as a consequence of having visually judged risk suggests that people (in this case patient with hemianopia) adjust their trajectory of reach, in order to avoid the risk of knocking over flankers in a target zone (Hesse, Lane, Aimola, & Schenk, 2012). Therefore we can see that risk affects the way we visually behave, and also may impact the physical element of our actions, investigating which factors related to the properties of objects and environments exhibit this effect on our visual treatment then would help us have better understanding of the high-level information that we extract from vision and use to guide behaviour.

The crucial part of these results that should be pursued in future work is the impact that these putdowns have on the overall temporal nature of the task and the microstructure of subsequent ORAs. Part of the main effort in the study of vision for action has been to explore the spatiotemporal allocation of gaze, one of the things that appears to impact that is the visual guidance of an object during the set down phase. The time between the eyes leaving one object after manipulation and arriving at another object, impacts the entire pace of the task as a whole. The properties of objects appear to influence the allocation of visual guidance when being put down on a surface, thereby impacting the point in time where the eyes are free to fixate the next object for manipulation. Further understanding of these factors that affect the temporal nature of our vision is vital if we are to understand the allocation of gaze during an active task.

8.7 Conclusions

We have demonstrated that various factors influence visual behaviour during an active task, the task itself, the objects used in the task, and whether we are familiar with the environment and objects. Our results have shown that visual behaviour changes as familiarity is acquired and that during this process we seem to be encoding something about the locations of even task irrelevant objects. Our visual behaviour appears to be modulated by the task we are completing, more specifically, by the objects involved in the task. The amount of time that we need to get our eyes to an object before the hand appears to change both as a result of the object and how familiar we are with it and the environment it is placed in. Similarly the amount of visual guidance we dedicate to the putting down of an object changes depending on the properties of an object.

The overall reduction in time for completing the same task in a novel vs. familiar task suggests that we get faster at even a familiar task, this however is not accounted for simply by a reduction in search time (Land et al., 1999) rather several elements of behaviour and visual behaviour change and speed up. In familiar environments the way we visually behave is different from unfamiliar or novel environments, even when completing a well-learned task. We spend less time visually searching, more time making looks ahead fixations and the microstructural elements of ORAs even appear to get faster. This findings in this thesis agree with Land et al. (1999) who point out if one is familiar with the environment then less conscious involvement is required to complete a task and the speed and ease with which objects are located is far less than in a novel environment, since presumably in novel environments there would be no opportunity to rely on exact representations of locations of objects. However we add to the current understanding by demonstrating that even the microstructure of the ORA is affected by the level of familiarity one has with the environment and objects.

We propose that the results of this thesis suggest that part of what is built up during the acquisition of familiarity is a representation of the environment and the objects, and that depending on the level of familiarity one has, the relative weighting of using visual information (i.e. performing visual search) vs. memory to locate objects will be affected. The level of detail that is built up during this period suggests that for objects that have been interacted with, specific information about their properties are encoded and consequently have an effect on the microstructural elements of ORAs involving them, but that for objects that were present in the room but task irrelevant, the level of encoding appears to only be in relation to the objects location in the environment, since we see a facilitatory effect on search behaviour but no changes in the elements of ORAs when the objects later have to be interacted with.

Furthermore, properties of the environment and objects appear to influence visual behaviour. The layout of the environment appears to affect the weighting of relying on memory to vision. Familiar environments that feature objects not in plain view without physical search appear to incur a stronger weighting on memory, with less time spent visually exploring the environment and less looks to task relevant objects, suggesting that people are relying more on representations, perhaps in an aid to avoid physical search. Properties of objects appear to influence the way we visually treat them during manipulations, certain objects, ad even objects in certain states (for example full or empty glasses in Chapter 4), appear to be fixated earlier and in receipt of more visual guidance during the end of an ORA, based on our results we would speculate that this may be as a result of the perceived risk of the object.

In conclusion we have demonstrated the flexibility of the way we visually behave depending on several circumstances, 1) the task we are undertaking 2) the properties of the objects we are manipulating, 3) the level of familiarity we have with the environment and objects, 4) the acquisition of familiarity 5) dealing with a task switch involving objects that had previously been visually available but irrelevant. We are flexible in terms of how we acquire information in these circumstances; how we use the information we already have to guide our vision, the temporal allocation of gaze and the speed of our subsequent manipulations. The findings contribute to our overall understanding of vision for action and further our knowledge of the types of factors we experience every day that affect the way we behave both generally and visually. We conclude that visual behaviour during an active task is flexible and able to respond appropriately to surroundings, circumstances and differing levels of prior knowledge in order to support our interactions with the world.

Chapter 9 References

Ackerley, R., & Barnes, G. R. (2011). The interaction of visual, vestibular and extra-retinal mechanisms in the control of head and gaze during head-free pursuit. *The Journal of Physiology*, *589*, 1627–1642. http://doi.org/10.1113/jphysiol.2010.199471

Aguiar, A., & Baillargeon, R. (2002). Developments in young infants' reasoning about occluded objects. *Cognitive Psychology*, *45*, 267–336. http://doi.org/10.1016/S0010-0285(02)00005-1

Ahissar, M., & Hochstein, S. (1993). Attentional control of early perceptual learning. *Proceedings of the National Academy of Sciences*, *90*(12), 5718–5722. http://doi.org/10.1073/pnas.90.12.5718

Ahrens, A. (1891). *Untersuchungen über die Bewegung der Augen beim Schreiben*. (C. Boldt, Ed.).

Aivar, M. P., Hayhoe, M. M., Chizk, C. L., & Mruczek, R. E. B. (2005). Spatial memory and saccadic targeting in a natural task. *Journal of Vision*, *5*(3), 177–93. http://doi.org/10.1167/5.3.3

Andrews, T. J., & Coppola, D. M. (1999). Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments. *Vision Research*, *39*, 2947–2953. http://doi.org/10.1016/S0042-6989(99)00019-X

Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, *17*(6-7), 1185–1204. http://doi.org/10.1080/13506280902978477

Ballard, D. H., Hayhoe, M. M., Li, F., & Whitehead, S. D. (1992). Hand-eye coordination during sequential tasks. *Philosophical Transactions of the*

Royal Society of London. Series B, Biological Sciences, *337*(1281), 331–8; discussion 338–9. http://doi.org/10.1098/rstb.1992.0111

Ballard, D. H., Hayhoe, M. M., & Pelz, J. B. (1995). Memory Representations in Natural Tasks. *Journal of Cognitive Neuroscience*, *7*, 66–80. http://doi.org/10.1162/jocn.1995.7.1.66

Bar, M. (2004). Visual objects in context. *Nature Reviews. Neuroscience*, *5*, 617–629. http://doi.org/10.1038/nrn1476

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*(1), 1–48. http://doi.org/10.18637/jss.v067.i01

Batista, A. P. (1999). Reach Plans in Eye-Centered Coordinates. *Science*, *285*(5425), 257–260. http://doi.org/10.1126/science.285.5425.257

Bekkering, H., & Neggers, S. F. W. (2002). VISUAL SEARCH IS MODULATED BY ACTION INTENTIONS. *Psychological Scienc*, *13*(4), 370–374. Retrieved from http://pss.sagepub.com/content/13/4/370.full.pdf

Belardinelli, A., Herbort, O., & Butz, M. V. (2015). Goal-oriented gaze strategies afforded by object interaction. *Vision Research*, *106*, 47–57. http://doi.org/10.1016/j.visres.2014.11.003

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scences. *Journal of Experimental Psychology*, *97*(1), 22–7. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/4704195

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology*, *14*, 143–177. http://doi.org/10.1016/0010-0285(82)90007-X

Biederman_1974_JEP_Scene_On_Info.pdf. (n.d.).

Brady, T. F., Konkle, T., Alvarez, G. a, & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(38), 14325–14329. http://doi.org/10.1073/pnas.0803390105

Briscoe, R., & Schwenkler, J. (2015). Conscious Vision in Action. *Cognitive Science*, *39*(7), 1435–67. http://doi.org/10.1111/cogs.12226

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.

Brockmole, J. R., & Henderson, J. M. (2006). Recognition and attention guidance during contextual cueing in real-world scenes: evidence from eye movements. *Quarterly Journal of Experimental Psychology (2006)*, *59*(7), 1177–1187. http://doi.org/10.1080/17470210600665996

Brouwer, A. M., & Knill, D. C. (2007). The role of memory in visually guided reaching. *Journal of Vision*, *7*(5), 6–6. http://doi.org/10.1167/7.5.6

Calvo-Merino, B., Grèzes, J., Glaser, D. E., Passingham, R. E., & Haggard, P. (2006). Seeing or Doing? Influence of Visual and Motor Familiarity in Action Observation. *Current Biology*, *16*, 1905–1910. http://doi.org/10.1016/j.cub.2006.07.065

Canosa, R. L. (2009). Real-world vision. *ACM Transactions on Applied Perception*, *6*(2), 1–34. http://doi.org/10.1145/1498700.1498705

Carpenter, R. H., & Williams, M. L. (1995). Neural computation of log likelihood in control of saccadic eye movements. *Nature*. http://doi.org/10.1038/377059a0

Castelhano, M., & Henderson, J. (2005a). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(March 2015), 1017–1040. http://doi.org/10.1080/13506280444000634

Castelhano, M., & Henderson, J. (2005b). Incidental visual memory for objects in scenes. *Visual Cognition*, *12*(6), 1017–1040. http://doi.org/10.1080/13506280444000634

Castelhano, M., Mack, M., & Henderson, J. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*. Retrieved from http://www.journalofvision.org/content/9/3/6.abstract

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior*, *18*(3), 257–273. http://doi.org/10.1016/S0022-5371(79)90146-4

Chun, M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0010028598906818

Chun, M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*. Retrieved from http://pss.sagepub.com/content/10/4/360.short

Chun, M. M. (2003). *Cognitive Vision*. *Psychology of Learning and Motivation* (Vol. 42). Elsevier. http://doi.org/10.1016/S0079-7421(03)01003-X

Cinelli, M. E., Patla, A. E., & Allard, F. (2009). Behaviour and gaze analyses during a goal-directed locomotor task. *Quarterly Journal of Experimental Psychology (2006)*, *62*(3), 483–99. http://doi.org/10.1080/17470210802168583

Cole, K. J. (2008). Lifting a familiar object: Visual size analysis, not memory for object weight, scales lift force. *Experimental Brain Research*, *188*(4), 551–

557. http://doi.org/10.1007/s00221-008-1392-y

Cooper, R., & Shallice, T. (2000). CONTENTION SCHEDULING AND THE CONTROL OF ROUTINE ACTIVITIES. *Cognitive Neuropsychology*. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/026432900380427

CROSSMAN, E. R. F. W. (1959). A THEORY OF THE ACQUISITION OF SPEED-SKILL∗. *Ergonomics*, *2*(2), 153–166. http://doi.org/10.1080/00140135908930419

Crundall, D., & Underwood, G. (2008). Some practical constraints on Cognitive Ethology: striking the balance between a theoretical approach and a practical methodology. *British Journal of Psychology (London, England : 1953)*, *99*, 341–345; discussion 355–359. http://doi.org/10.1348/000712608X283788

De Graef, P., Christiaens, D., & D’Ydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychological Research*, *52*, 317–329.

Delabarre, E. B. (1898). A method of recording eye movements. *The American Journal of Psychology*, *9*(4), 458–461.

Desanghere, L., & Marotta, J. J. (2011). “Graspability” of objects affects gaze patterns during perception and action tasks. *Experimental Brain Research*, *212*(2), 177–87. http://doi.org/10.1007/s00221-011-2716-x

Diaz, G., Cooper, J., Rothkopf, C., & Hayhoe, M. (2013). Saccades to future ball location reveal memory-based prediction in a virtual-reality interception task. *Journal of Vision*, *13*(1), 20–20. http://doi.org/10.1167/13.1.20

Dicks, M., Button, C., & Davids, K. (2010). Examination of gaze behaviors under in situ and video simulation task constraints reveals differences in information pickup for perception and action. *Attention, Perception, & Psychophysics*, *73*(3), 706–720.

Dodge, R. (1931). *Conditons and consequences of human variability*. New Haven: Yale University Press.

Dodge, R., & Cline, T. S. (1901). The angle velocity of eye movements. *Psychological Review*, *8*, 145–157.

Dodge, R., Travis, R. c., & Fox, J. C. (1930). Optic nystagmus III. Characteristics of the slow phase. *Archives of Neurology and Psychiatry*, *24*, 21–34.

Draschkow, D., Wolfe, J. M., & Võ, M. L. H. (2014). Seek and you shall remember: scene semantics interact with visual search to build better memories. *Journal of Vision*, *14*(8), 10. http://doi.org/10.1167/14.8.10

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–80. http://doi.org/10.1111/j.1467-9280.2006.01815.x

Eggert, T. (2007). *Eye Movement Recordings: Methods*. (A. Straube & U. Buttner, Eds.)*Neuro-opthamology* (Vol. 40). Basel: Krager. http://doi.org/10.1159/000100352

Ehinger, K. a., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(February 2015), 945–978. http://doi.org/10.1080/13506280902834720

Ehinger, K. A., Xiao, J., Torralba, A., & Oliva, A. (2011, July 1). Estimating scene typicality from human ratings and image features. Cognitive Science Society, Inc. Retrieved from http://dspace.mit.edu/handle/1721.1/71190

Epelboim, J., Steinman, R. M., Kowler, E., Edwards, M., Pizlo, Z., Erkelens, C.

J., & Collewijn, H. (1995). The function of visual search and memory in sequential looking tasks. *Vision Research*, *35*(23), 3401–3422. http://doi.org/10.1016/0042-6989(95)00080-X

Epelboim, J., Steinman, R. M., Kowler, E., Pizlo, Z., Erkelens, C. J., & Collewijn, H. (1997). Gaze-shift dynamics in two kinds of sequential looking tasks. *Vision Research*, *37*(18), 2597–2607. http://doi.org/10.1016/S0042-6989(97)00075-8

Fahle, M., Edelman, S., & Poggio, T. (1995). Fast perceptual learning in hyperacuity. *Vision Research*, *35*(21), 3003–3013. http://doi.org/10.1016/0042-6989(95)00044-Z

Findlay, J. F., & Gilchrist, I. D. (2003). *Active Vision. The Psychology of Looking and Seeing*. Oxford: Oxford University Press.

Flanagan, J. R., Bowman, M. C., & Johansson, R. S. (2006). Control strategies in object manipulation tasks. *Current Opinion in Neurobiology* , *16*(6), 650–659.

Flanagan, J. R., King, S., Wolpert, D. M., & Johansson, R. S. (2001). Sensorimotor prediction and memory in object manipulation.

Flanagan, J. R., & Wing, A. M. (1997). The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *17*(4), 1519–28. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9006993

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011a). Saccadic eye movements in a high-speed bimanual stacking task: changes of attentional control during learning and automatization. *Journal of Vision*, *11*, 9. http://doi.org/10.1167/11.7.9

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2011b). Saccadic eye movements in a high-speed bimanual stacking task: changes of attentional control during learning and automatization. *Journal of Vision*, *11*(7), 9. http://doi.org/10.1167/11.7.9

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2012a). Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task. *Journal of Vision*, *12*, 8–8. http://doi.org/10.1167/12.2.8

Foerster, R. M., Carbone, E., Koesling, H., & Schneider, W. X. (2012b). Saccadic eye movements in the dark while performing an automatized sequential high-speed sensorimotor task. *Journal of Vision*.

Foerster, R. M., & Schneider, W. X. (2013). Functionally sequenced scanpath similarity method (FuncSim): Comparing and evaluating scanpath similarity based on a task's inherent sequence of functional (action) units. *Journal of Eye Movement Research*, *6*, 1–22.

Foulsham, T., & Kingstone, A. (2013). Optimal and preferred eye landing positions in objects and scenes. *Quarterly Journal of Experimental Psychology (2006)*, *66*(9), 1707–28. http://doi.org/10.1080/17470218.2012.762798

Foulsham, T., Walker, E., & Kingstone, A. (2011). The where, what and when of gaze allocation in the lab and the natural environment. *Vision Research*, *51*(17), 1920–1931. http://doi.org/10.1016/j.visres.2011.07.002

Friedburg, C., Allen, C. P., Mason, P. J., & Lamb, T. D. (2004). Contribution of cone photoreceptors and post-receptoral mechanisms to the human photopic electroretinogram. *The Journal of Physiology*, *556*, 819–834. http://doi.org/10.1113/jphysiol.2004.061523

Friedman, A. (1979). Framing pictures: the role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology. General*, *108*(3), 316–55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/528908

Furmanski, C. S., & Engel, S. A. (2000). Perceptual learning in object recognition: object specificity and size invariance. *Vision Research*, *40*(5), 473–484. http://doi.org/10.1016/S0042-6989(99)00134-0

Gentilucci, M. (2002). Object motor representation and reaching – grasping control, *40*(May 2001), 1139–1153.

Gentilucci, M., Castiello, U., Corradini, M. L., Scarpa, M., Umiltà, C., & Rizzolatti, G. (1991). Influence of different types of grasping on the transport component of prehension movements. *Neuropsychologia*, *29*(5), 361–78. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/1886680

Gilchrist, I. D., North, A., & Hood, B. (2001). Is visual search really like foraging? *Perception*, *30*(12), 1459–1464. http://doi.org/10.1068/p3249

Glaser, R. (1987). Thoughts on Expertise. In C. Schooler & W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 81–94). NJ: Ablex: Norwood.

Golomb, J. D., Kupitz, C. N., & Thiemann, C. T. (2014). The influence of object location on identity: a “spatial congruency bias”. *Journal of Experimental Psychology. General*, *143*(6), 2262–78. http://doi.org/10.1037/xge0000017

Goodale, M. a, Meenan, J. P., Bülthoff, H. H., Nicolle, D. a, Murphy, K. J., & Racicot, C. I. (1994). Separate neural pathways for the visual analysis of object shape in perception and prehension. *Current Biology : CB*, *4*(7), 604–610. http://doi.org/10.1016/S0960-9822(00)00132-9

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between percieving objects and grasping them. *Nature*, *349*, 154–156.

Goodale, M. a., & Westwood, D. a. (2004). An evolving view of duplex vision: Separate but interacting cortical pathways for perception and action. *Current Opinion in Neurobiology*, *14*, 203–211. http://doi.org/10.1016/j.conb.2004.03.002

Guitteridge, M. (1939). *A study of motor achievements of young children*. Columbia University.

Guitton, D., & Volle, M. (1987). Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range. *Journal of Neurophysiology*, *58*(3), 427–459.

Hartley, T., Maguire, E. A., Spiers, H. J., & Burgess, N. (2003). The Well-Worn Route and the Path Less Traveled. *Neuron*, *37*(5), 877–888. http://doi.org/10.1016/S0896-6273(03)00095-3

Hayhoe, M. (2000). Vision Using Routines: A Functional Account of Vision. *Visual Cognition*, *7*(1997), 43–64. http://doi.org/10.1080/135062800394676

Hayhoe, M. M., McKinney, T., Chajka, K., & Pelz, J. B. (2012). Predictive eye movements in natural vision. *Experimental Brain Research*, *217*, 125–136. http://doi.org/10.1007/s00221-011-2979-2

Hayhoe, M. M., Mennie, N., Gorgos, K., Semrau, J., & Sullivan, B. (2004). The role of prediction in catching balls. *Journal of Vision*, *4*(2004), 156–156. http://doi.org/10.1167/4.8.156

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003a). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63. http://doi.org/10:1167/3.1.6

Hayhoe, M. M., Shrivastava, A., Mruczek, R., & Pelz, J. B. (2003b). Visual memory and motor planning in a natural task. *Journal of Vision*, *3*(1), 49–63. http://doi.org/10:1167/3.1.6

Henderson, J., & Hollingworth, A. (2003a). Eye movements and visual memory: Detecting changes to saccade targets in scenes. *Perception & Psychophysics*. Retrieved from http://link.springer.com/article/10.3758/BF03194783

Henderson, J., & Hollingworth, A. (2003b). Global transsaccadic change blindness during scene perception. *Psychological Science*. Retrieved from http://pss.sagepub.com/content/14/5/493.short

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. http://doi.org/10.1037/0278-7393.16.3.417

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*(February 2015), 1055–1082. http://doi.org/10.1080/13506280802685552

Henderson, J. M., Weeks, P. a. . J., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(1), 210–228. http://doi.org/10.1037/0096-1523.25.1.210

Herst, a. N., Epelboim, J., & Steinman, R. M. (2001). Temporal coordination of the human head and eye during a natural sequential tapping task. *Vision Research*, *41*, 3307–3319. http://doi.org/10.1016/S0042-6989(01)00158-4

Hespos, S. J., & Baillargeon, R. (2008). Young infants' actions reveal their developing knowledge of support variables: Converging evidence for violation-of-expectation findings. *Cognition*, *107*, 304–316. http://doi.org/10.1016/j.cognition.2007.07.009

Hesse, C., Lane, A. R., Aimola, L., & Schenk, T. (2012). Pathways involved in human conscious vision contribute to obstacle-avoidance behaviour. *The European Journal of Neuroscience*, *36*(3), 2383–90. http://doi.org/10.1111/j.1460-9568.2012.08131.x

Hidalgo-Sotelo, B., Oliva, A., & Torralba, A. (2005). Human Learning of Contextual Priors for Object Search: Where does the time go? In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops* (Vol. 3, pp. 86–86). IEEE. http://doi.org/10.1109/CVPR.2005.470

Hochberg, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary Theory and Research in Visual Perception* (pp. 309–331). New York: Holt.

Hoffman, R. R. (1996). How Can Expertise be Defined? Implications of Research from Cognitive Psychology. In R. Williams, W. Faulkner, & J. Fleck (Eds.), *Exploring Expertise* (pp. 81–100). Edinburgh: Edinburgh University Press.

Hollingworth, A. (2006). Visual memory for natural scenes: Evidence from change detection and visual search. *Visual Cognition*, *14*, 781–807. http://doi.org/10.1080/13506280500193818

Hollingworth, A. (2012a). Guidance of visual search by memory and knowledge. *Nebraska Symposium on Motivation. Nebraska Symposium on Motivation*, *59*, 63–89. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3875155&tool=pmcentrez&rendertype=abstract

Hollingworth, A. (2012b). Task specificity and the influence of memory on visual search: comment on Võ and Wolfe (2012). *Journal of Experimental Psychology. Human Perception and Performance*, *38*(6), 1596–603. http://doi.org/10.1037/a0030237

Hollingworth, A., & Henderson, J. M. (n.d.). Accurate visual memory for previously attended objects in natural scenes. Retrieved from http://psycnet.apa.orgjournals/xhp/28/1/113

Hollingworth, A., Williams, C. C., & Henderson, J. M. (2001). To see and remember: Visually specific information is retained in memory from previously attended objects in natural scenes. *Psychonomic Bulletin & Review*, *8*(4), 761–768. http://doi.org/10.3758/BF03196215

Hout, M. C., & Goldinger, S. D. (2010). Learning in repeated visual search. *Attention, Perception & Psychophysics*, *72*(5), 1267–82. http://doi.org/10.3758/APP.72.5.1267

Howard, C. J., Pharaon, R. G., Körner, C., Smith, A. D., & Gilchrist, I. D. (2011). Visual search in the real world: Evidence for the formation of distractor representations. *Perception*, *40*(10), 1143–1153. http://doi.org/10.1068/p7088

Huey, E. b. (1908). *The Psychology and Pedagogy of Reading: With a Review of the History of Reading and Writing and Methods, Texts and Hygiene in Reading*. The Macmillan Company.

Humphreys, G. W., & Forde, E. M. E. (1998). Disordered action schema and action disorganisation syndrome. *Cognitive Neuropsychology*, *15*((6)), 771–812.

Hunziker, H. W. (2006). *Im Auge des Lesers: vom Buchstabieren zur Lesefreude ; foveale und periphere Wahrnehmung*. Hunziker Hans-Werner. Retrieved from http://books.google.com/books?id=8EWbTwAPxrsC&pgis=1

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. http://doi.org/10.1016/S0042-6989(99)00163-7

Jakobson, L. S., Archibald, Y, M., Carey, D. P., & Goodale, M. A. (1991). A kinematic analysis of reaching and grasping movements in a patient recovering from optic ataxia. *Neuropsychologia*, *29*(8), 803–805.

Jeannerod, M. (n.d.). *The neural and behavioural organization of goal-directed movements. Oxford psychology series, No. 15.*

Jeannerod, M. (1981). Intersegmental coordination during reaching at natural visual objects. In *Long IJ, Baddeley A, eds. Attention and performance IX. Hillsdale: Lawrence Erlbaum* (pp. 153–169).

Jeannerod, M. (1986). Mechanisms of visuomotor coordination: A study in normal and brain-damaged subjects. *Neuropsychologia*, *24*, 41–78.

Jeannerod, M. (1999). Visuomotor channels: Their integration in goal-directed prehension. *Human Movement Science*, *18*(2-3), 201–218. http://doi.org/10.1016/S0167-9457(99)00008-1

Jiang, Y., & Wagner, L. C. (2004). What is learned in spatial contextual cuing— configuration or individual locations? *Perception & Psychophysics*, *66*(3), 454–463. http://doi.org/10.3758/BF03194893

Johansson, R. S. (1996). Sensory control of dexterous manipulation in humans.

Judd, T., Durand, F., & Torralba, A. (2012). A Benchmark of Computational Models of Saliency to Predict Human Fixations. *Mit-Csail-Tr-2012*, *1*. http://doi.org/1721.1/68590

Kanan, C., Tong, M. H., Zhang, L., & Cottrell, G. W. (2009). SUN: Top-down saliency using natural statistics. *Visual Cognition*, *17*(February 2015), 979–1003. http://doi.org/10.1080/13506280902771138

Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Three responses to Cognitive Ethology. *British Journal of Psychology*, *99*, 355–359. http://doi.org/10.1348/000712608X318644

Kirtley, C., & Tatler, B. W. (2015). Priorities for representation: Task settings and object interaction both influence object memory. *Memory & Cognition*. http://doi.org/10.3758/s13421-015-0550-2

Kit, D., Katz, L., Sullivan, B., Snyder, K., Ballard, D., & Hayhoe, M. (2014). Eye movements, visual search and scene memory, in an immersive virtual environment. *PloS One*, *9*(4), e94362. http://doi.org/10.1371/journal.pone.0094362

Knudsen, E. I. (2007). Fundamental components of attention. *Annual Review of Neuroscience*, *30*, 57–78. http://doi.org/10.1146/annurev.neuro.30.051606.094256

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–7. http://doi.org/10.1038/nature02169

Körner, C., & Gilchrist, I. D. (2007). Finding a new target in an old display: Evidence for a memory recency effect in visual search. *Psychonomic Bulletin & Review*, *14*(5), 846–851. http://doi.org/10.3758/BF03194110

Körner, C., & Gilchrist, I. D. (2008). Memory processes in multiple-target visual search. *Psychological Research*, *72*(1), 99–105. http://doi.org/10.1007/s00426-006-0075-1

Kowler, E. (1989). Cognitive Expectations, Not Habits, Control Anticipatory Smooth Oculomotor Pursuit. *Vision Research*, *29*(9), 1049–1057.

Kowler, E. (1990). The role of visual and cognitive processes in the control of eye movement. *Reviews of Oculomotor Research*.

Kowler, E., Martinc, A. J., & Pavel, M. (1984). The Effect of Expectations on

Slow Oculomotor Control - IV. Anticipatory Smooth Eye Movements Depend on Prior Target Motions, *21*(3), 197–210.

Krupinski, E. A., Tillack, A. A., Richter, L., Henderson, J. T., Bhattacharyya, A. K., Scott, K. M., … Weinstein, R. S. (2006). Eye-movement study and human performance using telepathology virtual slides: implications for medical education and differences with experience. *Human Pathology*, *37*(12), 1543–56. http://doi.org/10.1016/j.humpath.2006.08.024

Land, M. F. (1992). Predictable eye-head coordination during driving. *Nature*, *359*, 318–320. http://doi.org/10.1109/ICSMC.1993.385060

Land, M. F. (2004). The coordination of rotations of the eyes, head and trunk in saccadic turns produced in natural situations. *Experimental Brain Research*, *159*, 151–160. http://doi.org/10.1007/s00221-004-1951-9

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *352*(Land 1996), 1231–1239. http://doi.org/10.1098/rstb.1997.0105

Land, M. F., & Hayhoe, M. (2001). In what ways do eye movements contribute to everyday activities? *Vision Research*, *41*, 3559–3565. http://doi.org/10.1016/S0042-6989(01)00102-X

Land, M. F., & Lee, D. N. (1994). Where we look when we steer. *Nature*, *369*, 742–744. http://doi.org/10.1038/369742a0

Land, M. F., & McLeod, P. (2000). From eye movements to actions: how batsmen hit the ball. *Nature Neuroscience*, *3*(12), 1340–1345. http://doi.org/10.1038/81887

Land, M. F., & Tatler, B. W. (2001). Steering with the head: The visual strategy of a racing driver. *Current Biology*, *11*(Figure 3), 1215–1220.

http://doi.org/10.1016/S0960-9822(01)00351-7

Land, M. F., & Tatler, B. W. (2009). *Looking and Acting. Vision and Eye Movements in Natural Behaviour*. Oxford: Oxford University Press.

Land, M., Mennie, N., & Rusted, J. (1999). The roles of vision and eye movements in the control of activities of daily living. *Perception*, *28*(1992), 1311–1328. http://doi.org/10.1068/p2935

LandHorwood1995.pdf. (n.d.).

Levin, D. T., & Simons, D. J. (1997). Failure to detect changes to attended objects in motion pictures. *Psychonomic Bulletin & Review*, *4*(4), 501–506. http://doi.org/10.3758/BF03214339

Luchins, A. (1942). Mechanization in problem solving. *Psychological Monograph*, *248*(54), 6. http://doi.org/10.1037/h0093502

Mack, A., Rock, I., & Press, M. I. T. (1999). Etcetera. *Trends in Cognitive Sciences*, *3*(1), 13339.

Mack, A., Tang, B., Regina, T., & Kahn, S. (1992). Perceptual Organization and Attention.pdf. *Cognitive Psychology*, *24*, 475–501. http://doi.org/10.1093/schbul/sbr052

MACKWORTH, N. H., & THOMAS, E. L. (1962). Head-Mounted Eye-Marker Camera. *Journal of the Optical Society of America*, *52*(6), 713. http://doi.org/10.1364/JOSA.52.000713

Marteniuk, R. G., Leavitt, J. L., MacKenzie, C. L., & Athenes, S. (1990). Functional relationships between grasp and transport components in a prehension task. *Human Movement Science*, *9*(2), 149–176. http://doi.org/10.1016/0167-9457(90)90025-9

McClenaghan, B. A., & Gallahue, D. L. (1978). *Fundamental Movement: A developmental and remidial approach*. Saunders.

McKinney, T., Chajka, K., & Hayhoe, M. (2008). Pro-active gaze control in squash. *Journal of Vision* , *8* (6), 111. http://doi.org/10.1167/8.6.111

McLeod, P. (1987). Visual reaction time and high-speed ball games. *Perception*, *16*(1), 49–59. http://doi.org/10.1068/p160049

Mennie, N., Hayhoe, M., & Sullivan, B. (2007). Look-ahead fixations: anticipatory eye movements in natural tasks. *Experimental Brain Research*, *179*(3), 427–42. http://doi.org/10.1007/s00221-006-0804-0

Milner, A. D., & Goodale, M. A. (2006). *The Visual Brain in Action* (2nd ed.). Oxford: Oxford University Press.

Minsky, M. (1975). *The Psychology of Computer Vision*. (P. H. Winston, Ed.). New York: McGraw-Hill.

Mitroff, S. R., Simons, D. J., & Levin, D. T. (2004). Nothing compares 2 views: Change blindness can occur despite preserved access to the changed information. *Perception & Psychophysics*, *66*(8), 1268–1281. http://doi.org/10.3758/BF03194997

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: evidence for parallel programming of saccades. *Journal of Experimental Psychology. Human Perception and Performance*, *10*(5), 667–682. http://doi.org/10.1037/0096-1523.10.5.667

Murphy, G. L., & Wright, J. C. (1984). Changes in conceptual structure with expertise: Differences between real-world experts and novices. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *10*(1), 144–155. http://doi.org/10.1037/0278-7393.10.1.144

Myles-Worsley, M., Johnston, W. a, & Simons, M. a. (1988). The influence of expertise on X-ray image processing. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(3), 553–557.

http://doi.org/10.1037/0278-7393.14.3.553

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–21. http://doi.org/10.1016/j.visres.2005.08.025

Newby, E. A., & Rock, I. (1998). Inattentional blindness as a function of proximity to the focus of attention. *Perception*, *27*(9), 1025–1040. http://doi.org/10.1068/p271025

Newell, A., & Rosenbloom, P. S. (1993). Mechanisms of skill acquisition and the law of practice, 81–135. Retrieved from http://dl.acm.org/citation.cfm?id=162580.162586

Nickerson, R. S. (1965). AD6ZKU. *Canadian Journal of Psychology*, *19*(2), 155–160.

Nodine, C. F., Kundel, H. L., Lauver, S. C., & Toto, L. C. (1996). <title>Nature of expertise in searching mammograms for breast masses</title>. In H. L. Kundel (Ed.), *Medical Imaging 1996* (pp. 89–94). International Society for Optics and Photonics. http://doi.org/10.1117/12.236864

Norman, D. A., & Shallice, T. (1986). *Attention to action*. Springer.

Norman, G. R. (1989). The Development of Expertise in Dermatology. *Archives of Dermatology*, *125*(8), 1063. http://doi.org/10.1001/archderm.1989.01670200039005

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science (New York, N.Y.)*, *171*, 308–311. http://doi.org/10.1126/science.171.3968.308

Nuthmann, A., Smith, T. J., Engbert, R., & Henderson, J. M. (2010). CRISP: a computational model of fixation durations in scene viewing. *Psychological Review*, *117*(2), 382–405. http://doi.org/10.1167/8.6.1107

O’Regan, J. K. (n.d.). Solving the “real” mysteries of visual perception: The world as an outside memory. Retrieved from http://psycnet.apa.orgjournals/cep/46/3/461

O’Regan, J. K. (1992). Solving the “real” mysteries of visual perception: the world as an outside memory. *Canadian Journal of Psychology*, *46*(3), 461–488. http://doi.org/10.1037/h0084327

O’Regan, J. K., Rensink, R. A., & Clark, J. J. (1999). Change-blindness as a result of “mudsplashes”. *Nature*, *398*(6722), 34. http://doi.org/10.1038/17953

Oliva, A., Wolfe, J. M., & Arsenio, H. C. (2004). Panoramic search: the interaction of memory and vision in search through a familiar scene. *Journal of Experimental Psychology. Human Perception and Performance*, *30*(6), 1132–1146. http://doi.org/10.1037/0096-1523.30.6.1132

Olson, I. R., & Chun, M. M. (2002). Perceptual constraints on implicit learning of spatial context. *Visual Cognition*, *9*(3), 273–302. http://doi.org/10.1080/13506280042000162

Otto-De Haart, E. G., Carey, D. P., & Milne, A. B. (1999). More thoughts on perceiving and grasping the Muller-Lyer illusion. *Neuropsychologia*, *37*, 1437–1444. http://doi.org/10.1016/S0028-3932(99)00070-6

Over, E. a B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research*, *47*, 2272–2280. http://doi.org/10.1016/j.visres.2007.05.002

Palmer, T. E. (1975). The effects of contextual scenes on the identification of objects. *Memory & Cognition*, *3*(5), 519–526. http://doi.org/10.3758/BF03197524

Paulignan, Y., Frak, V. G., Toni, I., & Jeannerod, M. (1997). Influence of object position and size on human prehension movements. *Experimental Brain Research*, *114*(2), 226–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9166912

Pelz, J. B., & Canosa, R. (2001). Oculomotor behavior and perceptual strategies in complex tasks. *Vision Research*, *41*(25-26), 3587–3596. http://doi.org/10.1016/S0042-6989(01)00245-0

Perenin, M. T., & Vighetto, a. (1988). Optic ataxia: a specific disruption in visuomotor mechanisms. I. Different aspects of the deficit in reaching for objects. *Brain : A Journal of Neurology*, *111 (Pt 3*, 643–674. http://doi.org/10.1093/brain/111.3.643

Pezdek, K., Maki, R., Valencia-Laver, D., Whetstone, T., Stoeckert, J., & Dougherty, T. (1988). Picture memory: recognizing added and deleted details. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *14*(3), 468–476. http://doi.org/10.1037/0278-7393.14.3.468

Piaget, J. (1955). *The Child's Construction of Reality*. London: Routledge, Keegan Paul.

Potter, M. C., & Levy, E. I. (1969). Recognition memory for a rapid sequence of pictures. *Journal of Experimental Psychology*, *81*(1), 10–15.

Raj, R., Geisler, W. S., Frazor, R. A., & Bovik, A. C. (2005). Contrast statistics for foveated visual systems: fixation selection by minimizing contrast entropy. *Journal of the Optical Society of America A*, *22*(10), 2039. http://doi.org/10.1364/JOSAA.22.002039

Rayner, K. (2009). *Eye movements and attention in reading, scene perception, and visual search. Quarterly journal of experimental psychology (2006)* (Vol. 62). http://doi.org/10.1080/17470210902816461

Regan, D. (1992). Visual Judgments and misjudgments in cricket, and the art of the flight. *Perception*, *21*, 91–115.

Rensink, R. A. (2000). Seeing, sensing, and scrutinizing. *Vision Research*, *40*(10-12), 1469–1487. http://doi.org/10.1016/S0042-6989(00)00003-1

Rensink, R. A. (2002). Change detection. *Annual Review of Psychology*, *53*, 245–77. http://doi.org/10.1146/annurev.psych.53.100901.135125

Risko, E. F., Laidlaw, K., Freeth, M., Foulsham, T., & Kingstone, A. (2012). Social attention with real versus reel stimuli: toward an empirical approach to concerns about ecological validity. *Frontiers in Human Neuroscience*, *6*(May), 1–11. http://doi.org/10.3389/fnhum.2012.00143

Roach, N. W., Heron, J., & McGraw, P. V. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audio-visual integration. *Proceedings. Biological Sciences / The Royal Society*, *273*(1598), 2159–68. http://doi.org/10.1098/rspb.2006.3578

Roberton, M. A., & Halverson, L. E. (1984). *Developing children: Their changing movement*. Leas & Gebiger.

Rosman, B., & Ramamoorthy, S. (2011). Learning spatial relationships between objects. *The International Journal of Robotics Research*, *30*(11), 1328–1342. http://doi.org/10.1177/0278364911408155

Säfström, D., Johansson, R. S., & Flanagan, J. R. (2014). Gaze behavior when learning to link sequential action phases in a manual task. *Journal of Vision*, *14*(4), 3. http://doi.org/10.1167/14.4.3

Sailer, U., Flanagan, J. R., & Johansson, R. S. (2005). Eye-hand coordination during learning of a novel visuomotor task. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *25*(39), 8833–8842. http://doi.org/10.1523/JNEUROSCI.2658-05.2005

Schank, R. c. (1975). *Conceptual Information Processing*. New York: Elsevier.

Schütz-Bosbach, S., & Prinz, W. (2007). Perceptual resonance: action-induced modulation of perception. *Trends in Cognitive Sciences*, *11*(8), 349–355. http://doi.org/10.1016/j.tics.2007.06.005

Shackel, B. (1960). , Paris, France, Iliffe and Sons, London. In *In Proc. Second International Conf. on Medical Electronics*. London.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, *6*(1), 156–163. http://doi.org/10.1016/S0022-5371(67)80067-7

Shinoda, H., Hayhoe, M. M., & Shrivastava, A. (2001). What controls attention in natural environments? *Vision Research*, *41*, 3535–3545. http://doi.org/10.1016/S0042-6989(01)00199-7

Simons, D. J. (2000). Attentional capture and inattentional blindness. *Trends in Cognitive Sciences*, *4*(4), 147–155. http://doi.org/10.1016/S1364-6613(00)01455-8

Simons, D. J., & Chabris, C. F. (1999). Gorillas in our midst: Sustained inattentional blindness for dynamic events. *Perception*, *28*, 1059–1074. http://doi.org/10.1068/p2952

Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–7. http://doi.org/10.1016/S1364-6613(97)01080-2

Simons, D. J., & Rensink, R. A. (2005). Change blindness: past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20. http://doi.org/10.1016/j.tics.2004.11.006

Sims, C. R., Jacobs, R. A., & Knill, D. C. (2011). Adaptive Allocation of Vision under Competing Task Demands. *Journal of Neuroscience*, *31*(3), 928–943. http://doi.org/10.1523/JNEUROSCI.4240-10.2011

Sivak, M., Conn, L. S., & Olson, P. L. (1986). Driver eye fixations and the optimal locations for automobile brake lights. *Journal of Safety Research*, *17*(1), 13–22. http://doi.org/10.1016/0022-4375(86)90003-4

SIVAK, M., POST, D. V., OLSON, P. L., & DONOHUE, R. J. (1981). AUTOMOBILE REAR LIGHTS: EFFECTS OF THE NUMBER, MOUNTING HEIGHT, AND LATERAL POSITION ONREACTION TIMES OF FOLLOWING DRIVERS. *Perceptual and Motor Skills*, *52*(3), 795–802. http://doi.org/10.2466/pms.1981.52.3.795

Smith, A. D., Hood, B. M., & Gilchrist, I. D. (2008). Visual search and foraging compared in a large-scale search task. *Cognitive Processing*, *9*(2), 121–6. http://doi.org/10.1007/s10339-007-0200-0

Solman, G. J. F., & Smilek, D. (2010). Item-specific location memory in visual search. *Vision Research*, *50*(23), 2430–2438. http://doi.org/10.1016/j.visres.2010.09.008

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, *14*, 7. http://doi.org/10.1167/14.2.7

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2015). Disentangling the effects of spatial inconsistency of targets and distractors when searching in realistic scenes, *15*, 1–21. http://doi.org/10.1167/15.2.12.doi

Stainer, M. J., Anderson, A. J., & Denniss, J. (2015). Examination strategies of experienced and novice clinicians viewing the retina. *Ophthalmic & Physiological Optics : The Journal of the British College of Ophthalmic Opticians (Optometrists)*, *35*(4), 424–32. http://doi.org/10.1111/opo.12210

Standing, L. (1973). Learning 10,000 pictures. *The Quarterly Journal of Experimental Psychology*, *25*(2), 207–22. http://doi.org/10.1080/14640747308400340

Standing, L., Conezio, J., & Haber, R. N. (2013). Perception and memory for pictures: Single-trial learning of 2500 visual stimuli. *Psychonomic Science*, *19*(2), 73–74. http://doi.org/10.3758/BF03337426

Steinman, R. (2003). Gaze control under natural conditions. *The Visual Neurosciences*. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Gaze+Control+Under+Natural+Conditions#0

Sturz, B. R., Gaskin, K. A., & Roberts, J. E. (2014). Incidental encoding of enclosure geometry does not require visual input: evidence from blindfolded adults. *Memory & Cognition*, *42*(6), 935–42. http://doi.org/10.3758/s13421-014-0412-3

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining Priors and Noisy Visual Cues in a Rapid Pointing Task. *Journal of Neuroscience*, *26*(40), 10154–10163. http://doi.org/10.1523/JNEUROSCI.2779-06.2006

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*, *7*, 4.1–17. http://doi.org/10.1167/7.14.4

Tatler, B. W., Hirose, Y., Finnegan, S. K., Pievilainen, R., Kirtley, C., & Kennedy, A. (2013). Priorities for selection and representation in natural tasks. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *368*(1628), 20130066. http://doi.org/10.1098/rstb.2013.0066

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *366*, 596–610. http://doi.org/10.1098/rstb.2010.0188

Tatler, B. W., & Tatler, S. L. (2013a). The influence of instructions on object memory in a real-world setting. *Journal of Vision*, *13*, 5. http://doi.org/10.1167/13.2.5

Tatler, B. W., & Tatler, S. L. (2013b). The influence of instructions on object memory in a real-world setting. *Journal of Vision*, *13*(2), 5. http://doi.org/10.1167/13.2.5

Torok, N., Guillemin, V., & Barnothy, J. m. (1951). Photoelectric Nystagmography. *Ann Otol Rhinol Laryngol*, *60*, 917–927.

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. http://doi.org/10.1037/0033-295X.113.4.766

Underwood, G., Chapman, P., Bowden, K., & Crundall, D. (2002). Visual search while driving: skill and awareness during inspection of the scene. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 87–97. http://doi.org/10.1016/S1369-8478(02)00008-6

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003a). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*, 629–646. http://doi.org/10.1080/0014013031000090116

Underwood, G., Chapman, P., Brocklehurst, N., Underwood, J., & Crundall, D. (2003b). Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, *46*(6), 629–46. http://doi.org/10.1080/0014013031000090116

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency influence eye movements when inspecting pictures.

Quarterly Journal of Experimental Psychology (2006), *59*(February 2015), 1931–1949. http://doi.org/10.1080/17470210500416342

Underwood, G., Foulsham, T., & Humphrey, K. (2009). Saliency and scan patterns in the inspection of real-world scenes: Eye movements during encoding and recognition. *Visual Cognition*, *17*(6-7), 812–834. http://doi.org/10.1080/13506280902771278

Vaeyens, R., Lenoir, M., Williams, a M., Mazyn, L., & Philippaerts, R. M. (2007). The effects of task constraints on visual search behavior and decision-making skill in youth soccer players. *Journal of Sport & Exercise Psychology*, *29*(2), 147–169.

van Beers, R. J., Sittig, A. C., & Gon, J. J. D. van der. (1999). Integration of Proprioceptive and Visual Position-Information: An Experimentally Supported Model. *J Neurophysiol*, *81*(3), 1355–1364. Retrieved from http://jn.physiology.org/content/81/3/1355

Van Doorn, H., & Savelsbergh, G. J. P. (2007). *Grasping the Muller-Lyer illusion: the contribution of vision for perception and action*. *Neuropsychologia* (Vol. 45).

van Doorn, H., van der Kamp, J., de Wit, M., & Savelsbergh, G. J. P. (2009). Another look at the M??ller-Lyer illusion: Different gaze patterns in vision for action and perception. *Neuropsychologia*, *47*, 804–812. http://doi.org/10.1016/j.neuropsychologia.2008.12.003

Vincent, B. T., Baddeley, R., Correani, A., Troscianko, T., & Leonards, U. (2009). Do we look at lights? Using mixture modelling to distinguish between low- and high-level factors in natural image viewing. *Visual Cognition*, *17*(6-7), 856–879. http://doi.org/10.1080/13506280902916691

Võ, M. L.-H., & Henderson, J. M. (2009). Does gravity matter? Effects of semantic and syntactic inconsistencies on the allocation of attention during scene perception. *Journal of Vision*, *9*, 24.1–15. http://doi.org/10.1167/9.3.24

Võ, M. L.-H., & Schneider, W. X. (2010). A glimpse is not a glimpse: Differential processing of flashed scene previews leads to differential target search benefits. *Visual Cognition*, *18*(February 2015), 171–200. http://doi.org/10.1080/13506280802547901

Võ, M. L.-H., & Wolfe, J. M. (n.d.). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. Retrieved from http://psycnet.apa.orgjournals/xhp/38/1/23

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(1), 23–41. http://doi.org/10.1037/a0024147

Võ, M. L.-H., & Wolfe, J. M. (2012). When does repeated search in scenes involve memory? Looking at versus looking for objects in scenes. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(1), 23–41. http://doi.org/10.1037/a0024147

Wade, N. J., & Tatler, B. W. (2005). *The Moving Tablet of the Eye: The Origins of Modern Eye Movements*. Oxford: Oxford University Press.

Wang, S. H., Baillargeon, R., & Paterson, S. (2005). Detecting continuity violations in infancy: A new account and new evidence from covering and tube events. *Cognition*, *95*, 129–173. http://doi.org/10.1016/j.cognition.2002.11.001

Ward, P., & Williams, a M. (2003). Perceptual and Cognitive Skill Development in Soccer : The Multidimensional Nature of Expert Performance. *Journal of*

Sport & Exercise Psychology, *25*, 93–111.

Watts, R. G. (1991). Keep Your Eye On The Ball: The Science and Folklore of Baseball. *American Journal of Physics*, *59*(1991), 862. http://doi.org/10.1119/1.16719

Westheimer, G. (1954). Eye movement responses to a horizontally moving visual stimulus. *Archive of Opthmology*, *52*, 932–941.

Wilcox, T. (1999). Object individuation: Infants' use of shape, size, pattern, and color. *Cognition*, *72*, 125–166. http://doi.org/10.1016/S0010-0277(99)00035-9

Williams, A. M., Ford, P. R., Eccles, D. W., & Ward, P. (2011). Perceptual-cognitive expertise in sport and its acquisition: Implications for applied cognitive psychology. *Applied Cognitive Psychology*, *25*(3), 432–442. http://doi.org/10.1002/acp.1710

Williams, C. C., Henderson, J. M., & Zacks, f. (2005). Incidental visual memory for targets and distractors in visual search. *Perception & Psychophysics*, *67*(5), 816–827. http://doi.org/10.3758/BF03193535

Wohlschläger, A. (2000). Visual motion priming by invisible actions. *Vision Research*, *40*(8), 925–930. http://doi.org/10.1016/S0042-6989(99)00239-4

Wolfe, J. M. (1999). Inattentional amnesia. *Fleeting Memories: Cognition of Brief Visual Stimuli.*, 71–94. http://doi.org/10.1162/0898929053747685

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception & Psychophysics*, *73*(6), 1650–71. http://doi.org/10.3758/s13414-011-0153-3

Wolpert, D. ~M., Ghahramani, Z., & Flanagan, J. ~R. (2001). Perspective and Problems in Motor Learning. *Trends in Cognitive Sciences*, *5*(11), 487–494.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science-AAAS-Weekly Paper Edition*, *269(5232)*, 1880–1882.

Wu, C.-C., Wick, F. A., & Pomplun, M. (2014). Guidance of visual attention by semantic information in real-world scenes. *Frontiers in Psychology*, *5*, 54. http://doi.org/10.3389/fpsyg.2014.00054

Yang, S. N., & McConkie, G. W. (2001). Eye movements during reading: A theory of saccade initiation times. *Vision Research*, *41*, 3567–3585. http://doi.org/10.1016/S0042-6989(01)00025-6