Visual Attention and Stability

Sebastiaan Mathôt

VISUAL ATTENTION AND STABILITY

Members of the committee / *Leden van de commissie*

Dr. Artem Belopolsky
Prof. Dr. Heiner Deubel
Prof. Dr. Iain D. Gilchrist
Prof. Dr. Pieter Roelfsema
Prof. Dr. Jeroen Smeets

Nymphets / *Nymphanen*

Robert Boerwinkel, MSc
Theo Danes, BSc

VU UNIVERSITY AMSTERDAM

NWO
Netherlands Organisation for Scientific Research

VRIJE UNIVERSITEIT

Visual Attention and Stability

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Psychologie en Pedagogiek
op donderdag 24 januari 2013 om 11.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Sebastiaan Mathôt

geboren te Amsterdam

TABLE OF CONTENTS

∞

I. INTRODUCTION

Perceiving a stable world: A history of eye movements

Have you ever wondered why you move your eyes?

If you are like most people, eye movement researchers such as myself included, you have probably not. Perhaps the 'why' question so rarely comes to mind because the answer seems so obvious: We move our eyes to look at things! Of course it is true that we look at things, which requires movement of the eye. But this intuitive answer is incomplete, because it considers only those eye movements that are under voluntary control: *Saccades*, the rapid, shock-like eye movements that are so apparent when you look at another person's eyes (but not your own—try it with a mirror[1]). And *smooth pursuit*, the slow eye movements that allow you to track moving objects.

Although research has overwhelmingly focused on saccades and smooth pursuit (Kowler, 2011), three other types of eye movements are usually distinguished as well[2]: The *vestibulo-ocular reflex* (VOR), which compensates for motion of the head with a counter-rotation of the eye, the *optokinetic reflex* (OKR), which 'glues' our eyes to the visual environment, and *vergence*, which rotates the eyes inwards or outwards to focus on nearby or far-away objects respectively. These are gaze stabilizing reflexes that keep the eyes still with respect to the environment. The VOR does this by compensating for head movement, so that, for example, you can maintain eye contact when you shake your head. The OKR keeps your eyes fixed with respect to the visual environment, as you have no doubt experienced when looking out the window of a moving train or car. Vergence movements underlie gaze stabilisation as well, albeit less obviously, by maintaining a similarly centred retinal image in both eyes, thus allowing you to focus on objects at any distance.

Based on the observation that gaze stabilisation is an important function of eye movements, the evolutionary biologists Gordon Walls proposed his "one big idea, namely that the ancient and original function of the eye muscles was not really to move the eye but

---

1   The reason that you are unable to observe your own saccadic eye movements is that perception is almost completely suppressed (or 'omitted') during a saccade (Castet, 2010; Matin, 1974).

2   This conventional taxonomy does not include fixational eye movements, even though these may play a significant role in vision as well (e.g., Kuang, Poletti, Victor, & Rucci, 2012), nor accommodation.

rather to hold it still with respect to the environment" (Walls, 1962, p. 72). The evolutionary history of eye movements is therefore one of visual stability: Eye movements evolved to avoid retinal motion, to make sure that the retinal image remains stable during self-generated movement of the organism. "Otherwise", argues Walls, "he [the organism] would think things to be in motion which were not, and really-moving objects, enemies and foods, would be difficult or impossible to notice and deal with" (Walls, 1962, p. 71).

Although Walls' insights have not been widely adopted, they have been recognised and extended by Michael Land (1999; Land & Nilsson, 2002). According to Land, one reason why gaze stabilisation is crucial is that photoreceptors are sluggish, requiring approximately 20ms of constant stimulation to reach a saturated response (Land, 1999; see also Srinivasan & Bernard, 1975). If the image drifts across the retina, even with only a moderate speed, photoreceptors will not receive constant stimulation for a sufficient period of time. This results in a blurred visual percept, or, more specifically, in a reduced sensitivity for high spatial frequencies (Burr & Ross, 1982; Kelly, 1979).

In addition, gaze stabilisation facilitates motion detection. If self-generated retinal motion is eliminated, all retinal motion corresponds to real movement 'out there' and motion detection is trivial. Land (1999) takes this argument to an interesting extreme. It has been known for some time that perfect elimination of retinal motion results in a fading of the percept (Gerrits, De Haan, & Vendrik, 1966; Riggs, Ratliff, Cornsweet, & Cornsweet, 1953): The image appears to vanish, including the *eigengrau* that is normally perceived in darkness, and only moving objects remain visible[3]. Humans are not, under normal circumstances, able to achieve perfect retinal stabilisation of this type (e.g., Ferman, Collewijn, Jansen, & Van den Berg, 1987), but Land (1999) speculates that certain animals, notably prey animals such as rabbits and squirrels, might be able to do so (there is no direct evidence for this hypothesis, but see Martinez-Conde & Macknik, 2008). This could provide these animals with an efficient mechanism for detecting approaching predators, which would be among the few visible objects in an otherwise absent visual environment. This idea is interesting, because, in a sense, it posits an extreme form of automatic attentional capture by salient objects, as will be discussed in the next section (cf. Theeuwes, 2010; Yantis & Jonides, 1984).

A third reason for stabilising the retinal image, not described by Land (1999) but nevertheless crucial and the focus of the present thesis, is that it facilitates the maintenance

---

3   This may seem at odds with the claim that retinal motion is detrimental to perception, but it is not. Too much retinal motion is detrimental, yet some is required to prevent the image from fading (cf. Gerrits, De Haan, & Vendrik, 1966; Riggs, Ratliff, Cornsweet, & Cornsweet, 1953), and a slight fixational drift may even have some benefits with respects to image processing (Kuang et al., 2012).

of positional information. This is particularly true for tasks that involve some form of memory or attention. Consider, for example, a task in which participants remember the location of a briefly presented stimulus (e.g., Golomb & Kanwisher, 2012; see also Chapter v of this thesis). If the retinal image remains stable during the retention interval, the participant can simply rely on the retinal location of the (remembered) stimulus. This is easy, because the visual system is largely retinotopically organised and so the participant can rely on his or her 'native' frame of reference (Gardner, Merriam, Movshon, & Heeger, 2008; Golomb & Kanwisher, in press; Mathôt & Theeuwes, 2011a; Talbot & Marshall, 1941). However, if the retinal image has moved during the retention interval, for example because of eye movements, the resulting retinal displacement needs to be taken into account. This is not a trivial feat to accomplish, even for our finely tuned visuomotor system, and leads to reduced accuracy of positional memory (Golomb & Kanwisher, 2012; for related findings, see Bridgeman, Hendry, & Stark, 1975; Deubel, Schneider, & Bridgeman, 1996).

To summarise the argument put forward by Walls (1962) and Land (1999; Land & Nilsson, 2002), there is good reason to believe that eye movements evolved primarily as a means of gaze stabilisation. This illustrates the importance of a stable retinal image for visual perception. It is therefore striking that eye movements are also the source of the most violent disruptions of visual stability. For example, consider what happens if you look out the window of a moving train. Initially, your eyes will slowly rotate to stabilise the retinal image of the sliding landscape (the *slow phase* of the OKR). Yet this slow compensatory rotation cannot go on indefinitely, because the eyes quickly reach their maximum eccentricity. When this happens, the eyes snap back to a more comfortable position, with a fast, saccade-like eye movement (the *fast phase*). The retinal image motion that results from this snapping back is dramatic, reaching speeds of up to 100°/s (based on 4° saccades cf. Sparks, 2002). Gaze stabilising reflexes are thus characterised by a trade-off: A brief period of severe instability is traded against a longer period during which the retinal image is kept stable. Because the fast phase of the OKR is indeed very fast, there is only a brief moment during which the sluggishness of photoreceptors play a role and motion detection is impaired (Land, 1999; Land & Nilsson, 2002). Yet high velocities do not alleviate the localisation problem that is associated with displacement of the retinal image.

Therefore, eye movements are both a mechanism for, and a challenge to visual stability. In humans and other foveate animals, the challenge is particularly great, because we have adapted gaze stabilising reflexes in a way that is diametrically opposed to their original

function. The slow phase of the OKR forms the basis for voluntary smooth pursuit, whereas the fast phase has evolved into voluntary saccades (Robinson, 1968)[4]. Each saccadic and smooth pursuit eye movement results in a displacement of the projection of the world onto the retina—These eye movements disrupt rather than preserve a stable retinal imagine. Yet, even though visual perception is ultimately derived from the retinal image, we do not perceive any displacements or disruptions. And even more importantly, we are able to effortlessly execute visually guided movements, apparently unperturbed by the instability of visual input at the level of the retina.

The question of how we are able to function efficiently, despite the continuously shifting retinal image, is the subject of the present thesis. This is a broad question, which can be investigated from many angles, and I have focused in particular on the role of attention in visual stability. In the following sections of this introduction, I will first introduce the concept of visual attention. Next, I will outline a view on perception in general, and visual stability in particular, which has guided the questions that I have asked, and in which the experimental results described in this thesis can be understood. Finally, I will present an overview of the present thesis through a brief summary of each chapter.

<div align="center">

"Everyone knows"

</div>

As William James famously pointed out, "everyone knows what attention is" (James, 1890, p. 404). Attention is a faculty of mind that allows us to focus on some aspect of our environment in order to see, hear, feel, or otherwise experience that particular aspect more clearly. The focus of attention can be an object, a location, a sound, a feeling, a colour, or even a memory. Attention is ubiquitous.

Yet the fact that everybody (kind of) knows what attention is did not stop William James, nor the army of researchers that followed in his footsteps, from spending considerable time and effort disentangling and characterising this slippery phenomenon. A full review of the literature can be found in Chapter ii (for recent reviews, see also Carrasco, 2011; Theeuwes, 2010), but for the purpose of this introduction I will provide only a brief discussion of the most important findings. It is fair to say that research has largely confirmed the common sense view of attention: Some objects, or rather events, attract our attention automatically (e.g., Theeuwes, 1991; Yantis & Jonides, 1984). A flash of light, for example, or a previously unperceived soccer ball that you suddenly realise is on collision course with your head (Franconeri & Simons, 2003a). It is not hard to imagine

---

4   The notion that voluntary eye movements are essentially a form of encephalised, voluntary OKR is occasionally mentioned explicitly (Robinson, 1968) or hinted at (Easter, 1972; Land, 1999; Land & Nilsson, 2002; Walls, 1962). However, to the best of my knowledge it has never been directly investigated and thus remains speculative.

the purpose of this automatic orienting of attention: It allows organisms to respond reflexively to threatening events that require immediate action, such as suddenly approaching predators. Many of the studies described in this thesis deal with this type of automatic, exogenous attention. But humans are not purely driven by their environment. We are able to exert considerable, albeit imperfect volitional control over where we direct our attention (e.g., Posner, 1980). This gives you the ability to focus on this text, for example, even though it is not particularly salient. Furthermore, we can attend to an object without looking directly at it. Covertly attending to something is effortful, because it requires us to suppress the direct link that exists between attention and eye movements (e.g., Deubel & Schneider, 1996; for a review, see Craighero & Rizzolatti, 2005). Therefore, in real life gaze generally coincides with the focus of attention. Yet in the lab, attention is frequently studied in the absence of eye movements, an approach that you will encounter in the present thesis as well (e.g., Chapter vii).

You will not be surprised by the characterisation of visual attention outlined above. But some more counter-intuitive findings have emerged from attention research as well. One phenomenon that will figure prominently throughout this thesis is *change blindness* (Grimes, 1996; O'Regan, Deubel, Clark, & Rensink, 2000; Rensink, O'Regan, & Clark, 1997; Simons & Rensink, 2005; also see Mack & Rock, 1998; Mathôt & Theeuwes, 2012). Change blindness is as famous as it is profound, and there is a good chance that you have seen a demonstration of it. A frequently used technique to demonstrate change blindness is the flicker paradigm: An image, typically a photo of a natural scene, is presented. This is followed by a brief blank display. Next another image is presented. This second image is a copy of the first that has been modified in some obvious way. A large tree may have been edited out, for example. After the presentation of the second image another blank display is presented. This presentation sequence is repeated until the participant has identified the change. The crucial finding is that participants are very poor at detecting the change, and that this poor performance is entirely due to the seemingly innocuous blank display that separates the two images. If the two images are presented in immediate succession, participants spot the change right away. The difference in response times between the blank and no-blank conditions is enormous, on the order of a magnitude.

What do the results from change blindness experiments tell us? The typical interpretation is that the blank prevents the change from being a unique visual event, by having it coincide with the onset of the display. Therefore, our attention is no longer automatically, exogenously drawn towards the change. In itself, this is not too surprising, and some performance decrement might be expected. What is surprising, however, is the

size of the effect: The finding that we are *so extremely poor* at detecting changes without a unique visual event to guide us. Apparently, we are unable to the monitor the entire visual image at once, and fall back to scanning the image in a more or less serial fashion until, by chance, we happen to focus on the part of the scene that changes. This illustrates that we do not perceive our entire visual surroundings at once, although in a sense it feels like we do.

Change blindness shows that attention is much more than a subtle reallocation of perceptual resources, or a bias that allows us to perceive one thing a bit more clearly than the other. It is the *sine qua non* for all but the most basic forms of perception. We can therefore revisit the definition of attention with which we started this section:

Attention is a faculty of mind that allows us to focus on some aspect of our environment in order to see, hear, feel, or otherwise experience that particular aspect ~~more clearly~~.

A weak non-representationalist account of visual perception

Most of us, layman and scientists alike, subscribe to some form of *representationalism*, which is a fancy way of saying that we assume the existence of a cognitive representation of the environment. We assume that something inside our brain mirrors what is 'out there'. Philosophers enjoy making fun of the hypothetical, naïve representationalist who believes that our brain contains a blue substance that allows us to experience the colour blue. But this is silly and an unfair characterisation of the representationalist view. Most scientists take a much more abstract view. Just like few people believe that digital photos are stored in the form of little coloured pictures on the hard drive of a computer, few (if any) scientists believe that the brain contains something like an actual picture, even though some may speak, colloquially, of a 'picture in the brain'. The crux of representationalism is the assumption that it is possible, at least in theory, to reconstruct the contents of visual perception from brain activity, just like it is possible to reconstruct a photo from a digital file (for a recent perspective, see Kanai & Tsuchiya, 2012).

And this is not silly at all. In fact, since the advent of functional magnetic resonance imaging (fMRI) in the early 90's of the previous century, crude mind reading of this type has become commonplace. A striking demonstration of this is a study by Harrison and Tong (2009). In their study, participants briefly saw two patches of lines, each of a different orientation. The participants were instructed to keep one of these patches in memory. The crucial finding of Harrison and Tong (2009) was that they were able to predict which of the stimuli was being held in memory with more than 80% accuracy, based solely on brain activity from the interval when the stimuli were no longer visible. Clearly, the brain contained some kind of representation of the remembered stimulus,

which Harrison and Tong (2009) were able to decipher (see also Logothetis & Schall, 1989; Norman, Polyn, Detre, & Haxby, 2006; Pasley et al., 2012; Tong, Nakayama, Vaughan, & Kanwisher, 1998).

Visual stability poses a serious problem in the context of representationalism. The general consensus is that visual input is processed in a gaze-centred, or retinotopic, frame of reference (Gardner et al., 2008; Golomb & Kanwisher, in press; Talbot & Marshall, 1941; for recent reviews, see Mathôt & Theeuwes, 2011a; Wurtz, 2008). This means that with each eye movement the representation of our visual surroundings becomes misaligned with the world. Since we do not experience any such misalignment, it seems that there must be some mechanism that updates our visual representations to compensate for these retinal displacements. Much of the research on visual stability has focused on identifying such a mechanism, either by characterising it on a behavioural level (e.g., Golomb, Chun, & Mazer, 2008; Irwin, 1991; Mathôt & Theeuwes, 2010a, 2010b, 2012; Melcher, 2005; O'Regan & Lévy-Schoen, 1983; Rolfs, Jonikaitis, Deubel, & Cavanagh, 2011) or by identifying its neural substrate (e.g., Duhamel, Colby, & Goldberg, 1992; Kusunoki & Goldberg, 2003; Morris, Kubischik, Hoffmann, Krekelberg, & Bremmer, 2012; Sommer & Wurtz, 2006).

However, the search for an active mechanism for visual stability is driven by the assumption that there are retinotopically organised representations that require such a compensatory process—a premise which is by no means universally accepted. In an influential and provocative review, Kevin O'Regan and Alva Noë (2001; also see Noë & O'Regan, 2000; Noë, Pessoa, & Thompson, 2000; MacKay, 1972) proposed a *non-representationalist* account of visual perception. They, along with many philosophers, reject a central role for cognitive representations, and posit that "visual experience does not arise because an internal representation of the world is activated in some brain area" (O'Regan & Noë, 2001, p. 946). Instead, they argue that there is usually little need for cognitive representations, because the world is right there in front of our eyes, and thus serves as an 'external representation' or 'external memory' (O'Regan, 1992). According to this view, there is nothing paradoxical about the results from change blindness experiments (e.g., Rensink et al., 1997). The reason that we feel as though we see everything at once, even though change blindness shows that we do not, is that everything is out there to see. We know that we can inspect any part of our visual surroundings at will, by directing our attention towards it, and this knowledge gives us the feeling that the world is continuously present (which, of course, it is). This view is also supported by the many troubling (in a representationalist view) dissociations that are found in visual perception, such as illusions

that paradoxically affect perceived motion, but not position (Smeets, Brenner, de Grave, & Cuijpers, 2002): If there are no unitary representations, it is no surprise that an illusion can have different, even contradictory effects on different aspects of perception.

With respect to visual stability, non-representationalism has profound consequences for the questions that one can sensibly ask. In its most extreme form, it reduces visual stability to a non-issue, because "from this viewpoint, there is no need to postulate a mechanism that re-positions the retinal image after eye saccades so that the world appears stationary" (O'Regan & Noë, 2001, p. 950). After all, if there are no cognitive representations, then no compensatory process is required to make sure that these representations remain aligned with the world 'out there'.

In the course of my research, my own view has shifted gradually, but distinctly towards a weak form of non-representationalism. Consequently, you will find references to the conscious perception of a stable world and the 'problem' of visual stability rather more frequently in the earlier chapters of this thesis than in the later ones (as the order of the chapters is by and large chronological). The reason for this shift is the picture of trans-saccadic integration (our ability to 'fuse' visual information from one fixation to the next) that has emerged from research over the past decades (Irwin, 1996), and in particular over the past few years (Bays & Husain, 2007; Cavanagh, Hunt, Afraz, & Rolfs, 2010; Mathôt & Theeuwes, 2011a). Simply put, there have been no convincing demonstrations of trans-saccadic integration of localised[5], detailed, perceptual information—and not for lack of trying (e.g. Irwin, Yantis, & Jonides, 1983; McKyton, Pertzov, & Zohary, 2009; O'Regan & Lévy-Schoen, 1983). Although studies on adaptation aftereffects seemed promising in this regard (Ezzati, Golzar, & Afraz, 2008; Melcher, 2005, 2007, 2008a, 2008b; see also Wittenberg, Bremmer, & Wachtler, 2008), it has recently been disputed whether such findings indeed reflect reflect trans-saccadic integration (Afraz & Cavanagh, 2009; Hunt & Cavanagh, 2009; Knapen, Rolfs, & Cavanagh, 2009; Wenderoth & Wiese, 2008; also see Chapter vi of this thesis). There has been positive evidence for trans-saccadic integration, but only of abstract representations, conceptual rather than visual (e.g., Henderson & Siefert, 2001; Pollatsek, Rayner, & Collins, 1984). Therefore, pending strong evidence to the contrary, the conclusion must be that we do not, or hardly, integrate ('remap') visual representations from one fixation to the next (Bays & Husain,

---

5   Some studies have shown trans-saccadic priming using artificial, non-semantic stimuli (Demeyer, De Graef, Wagemans, & Verfaillie, 2009, 2010), or have described detailed verbal reports of the contents of fixation after an artificial 'blindfold' (Tatler, 2001). These results could potentially revive the notion of a limited (semi-)persistent low-level perceptual representation. However, these studies have not tested the extent to which these effects are location-specific, and the implications for visual stability are therefore not clear.

2007; Cavanagh et al., 2010; Irwin, 1996; O'Regan & Lévy-Schoen, 1983; O'Regan & Noë, 2001). Consequently, visual stability might indeed be a non-issue in the context of conscious visual perception.

Yet, like most things, non-representationalism is untenable when taken to its extreme. For one, there are many 'mind reading' studies, such as the one by Harrison and Tong (2009) discussed above, that unequivocally show that some neural representations exist (Logothetis & Schall, 1989; Norman et al., 2006; Pasley et al., 2012; Tong et al., 1998). Furthermore, the fact that we can memorise and attend to (locations of) stimuli that are no longer present means that some form of representation, however sparse, must exist. After all, we cannot rely on the world as an external representation when it comes to attention and related internal states, such as working memory and action preparation.

One of the conclusions of the present thesis is therefore that visual stability should be considered in the context of attention, rather than in the context of conscious visual perception. Unlike O'Regan and Noë (2001) have argued, there is good evidence, some of which presented in this thesis, for the existence of an active 'remapping' mechanism for visual stability. But this mechanism operates in the domain of attention, and its relevance to conscious visual perception is debatable (cf. Bays & Husain, 2007). In all likelihood, rich cognitive representations of the external world in all its detail and glory do not exist, and they are consequently not subject to trans-saccadic integration. But sparse representations, perhaps barely more than positional information, are.

Chapter overview

With the exception of Chapters v and vi, all chapters are directly based on published articles, with only minor textual corrections and updated references. Where necessary, footnotes have been added to comment on recent developments.

Part 1: Visual attention and stability

Chapter ii (Mathôt & Theeuwes, 2011a) is a literature review in which we describe two distinct mechanisms that underlie visual stability. The first mechanism is passive (hence perhaps more aptly labelled a 'phenomenon'): We assume that objects are stable across eye movements, unless there is considerable evidence to the contrary (Bridgeman et al., 1975; Currie, McConkie, Carlson-Radvansky, & Irwin, 2000; Deubel et al., 1996). The second is an active mechanism, often called *remapping* or *spatial updating*, that updates retinotopic representations to compensate for eye movements (Duhamel et al., 1992; Wurtz, 2008). Following the view outlined in this introduction and developed further in Chapters v and vi, remapping and the assumption of stability may reflect visual stability in

the context of attention (and action-preparation and working memory) and perception, respectively (Bays & Husain, 2007).

Chapter iii (Mathôt & Theeuwes, 2010a) describes three experiments that investigate the dynamics of attention in the interval surrounding an eye movement. In line with a series of studies by Golomb and colleagues (2008), we found in Experiment 1 that attention partly shifts along with a saccadic eye movement, at least transiently. That is, right after an eye movement, attention has both a retinotopic and a spatiotopic component. We interpreted this finding as evidence for partial, but imperfect maintenance of visual stability. Interestingly, in Experiments 2 and 3 we found that the locus of attention was also systematically affected before the onset of a saccadic eye movement, an effect that we attributed to pre-saccadic receptive field changes (cf. Duhamel et al., 1992). This interpretation has recently been criticised by Rolfs and colleagues (2011), who argued that one should not expect a shift of attention in the direction of an eye movement (cf. Biber & Ilg, 2011; W. Harrison, Remington, & Mattingley, 2012; Mathôt & Theeuwes, 2010a; Melcher, 2007), but rather in the direction opposite from the eye movement. As we describe in Chapter ii (see Figure ii.3), they make a valid point, although, contrary to their claim, the concept of predictive remapping allows one to predict a shift in either direction. It will be an interesting avenue for future research to explore precisely how the findings reported here and by Rolfs and colleagues (2011) are related to predictive remapping.

In **Chapter iv** (Mathôt & Theeuwes, 2010b) we follow up on Chapter iii by showing that inhibition of return (IOR) similarly has both a spatiotopic and retinotopic component. Crucially, we mapped out these components as a function of the post-saccadic delay (the interval between the end of one saccade, and the onset of the next), and found a distinctly different time course for both components. More specifically, immediately after an eye movement IOR was predominantly retinotopic, whereas at later post-saccadic intervals IOR was predominantly spatiotopic. We interpreted this finding as reflecting a gradual recovery of visual stability, or evidence for an active remapping process.

In **Chapter v** we describe a series of localisation experiments that investigated whether and when visual stability is preserved during smooth pursuit eye movements. Participants reported the location of a briefly presented target, while they tracked a smoothly moving dot with their eyes (cf. Mateeff, Mitrani, & Stojanova, 1982; Mateeff, Yakimoff, & Dimitrov, 1981). Crucially, in Experiment 1 participants localised the target by making a saccadic eye movement to its (remembered) location. In Experiments 2 and 3, participants localised the target by reporting whether a subsequently presented probe stimulus was presented to the left or to the right of the target. In all experiments, we investigated

whether the retinotopic bias (the tendency to mislocalise the target in the direction of the smooth pursuit eye movement) was persistent or transient, in the sense that it was present only for fast responses and short retention intervals. In line with the results described in Chapter iv, we found a transient retinotopic bias for the direct (saccadic) localisation task (Blohm, Missal, & Lefevre, 2005). However, for the indirect (probe matching) task, we found a substantial and persistent retinotopic bias. We interpreted these findings as evidence for a dissociation between visual stability for action and perception. More specifically, we proposed that visual stability is only maintained when participants actively engage with the location of a perceived stimulus (cf. Bays & Husain, 2007).

In **Chapter vi** we focus on the controversy surrounding the reference frame of adaptation aftereffects. Until recently, a widely held assumption was that adaptation to orientation, motion, and faces occurs at least partly in a spatiotopic frame of reference (Ezzati et al., 2008; Melcher, 2005, 2008a, 2008b; see also Wittenberg et al., 2008). This was generally taken to be strong evidence for remapping of low-level visual features across saccades. However, a more recent series of studies has shown an exclusively retinotopic frame of reference for the same adaptation aftereffects (Afraz & Cavanagh, 2009; Knapen et al., 2009; Knapen, Rolfs, Wexler, & Cavanagh, 2010; Wenderoth & Wiese, 2008). Here we sought to determine anew what the reference frame of the tilt-adaptation aftereffect (TAE) is. The results were clear-cut. In our hands at least, TAE is consistently and exclusively retinotopic. This is in line with an action-oriented view in which there is no need for remapping of low-level perceptual features (cf. Bays & Husain, 2007; Cavanagh et al., 2010; O'Regan & Noë, 2001).

In **Chapter vii** (Mathôt & Theeuwes, 2012) we consider blindness to changes across saccades. More specifically, we investigated whether a suddenly appearing stimulus captures attention when it is presented while the eyes are in motion. The crucial finding was that an intra-saccadically presented additional stimulus had no effect beyond that of increasing the set size by one item. The implications of this result are threefold: There is substantial blindness for intra-saccadic changes, even in sparse artificial displays (Grimes, 1996); Stimuli do not capture attention in the absence of a visual transient (Franconeri, Hollingworth, & Simons, 2005); Trans-saccadic integration is insensitive to large changes, such the appearance of an additional stimulus (Bridgeman et al., 1975).

In **Chapter viii** (Mathôt, Hickey, & Theeuwes, 2010) we focus on the properties of attention per se, rather than the interaction with visual stability. More specifically, we investigated the interference caused by a distractor stimulus as a function of whether target and distractor were presented in the same, or in opposite visual fields. The crucial finding

was that the polarity of the hemifield effect depended on whether or not the stimuli were presented simultaneously. If the target and distractor were presented (almost) simultaneously, interference was largest when they were presented in the same visual field, indicative of competitive interactions (Desimone & Duncan, 1995; Mounts, 2000a). However, if the distractor was presented prior to the target, interference was greatest when the stimuli were presented in opposite hemifields, reflecting a cost of reorienting attention across the vertical meridian (Rizzolatti, Riggio, Dascola, & Umiltá, 1987).

Part II: Behavioural methods

The second part of the present thesis describes a more pragmatic line of research, or rather development. During my PhD project we have developed a number of software packages and algorithms. Three of these have been described in methodological papers and made publicly available.
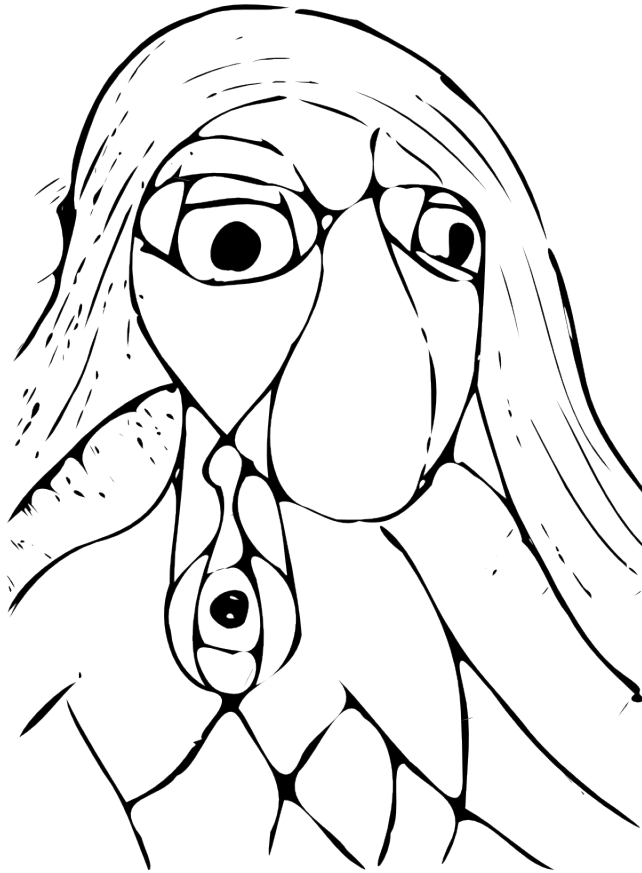
In **Chapter ix** (Mathôt, Schreij, & Theeuwes, 2012) we describe OpenSesame, a graphical experiment builder for the social sciences. We have developed OpenSesame as a free alternative to proprietary software packages such as E-Prime (Psychology Software Tools, Sharpsburg, PA, USA), Presentation (Neurobehavioral Systems, Albany, CA, USA), and Experiment Builder (SR Research, Mississauga, ON, Canada). Many of the studies described in this thesis have been conducted using OpenSesame.

In **Chapter x** (Mathôt & Theeuwes, 2011b) we describe Mantra, a webcam-based object/ movement tracker, designed specifically for use in psychological experiments. The algorithm behind Mantra, which relies on colour matching, is simple, but effective and highly sensitive.

In **Chapter xi** (Mathôt, Cristino, Gilchrist, & Theeuwes, 2012) we describe an algorithm to estimate the similarity between pairs of eye movement sequences. Similarity ratings are a powerful tool for analysing complex sets of eye movement data (e.g., Cristino, Mathôt, Theeuwes, & Gilchrist, 2010; Hacisalihzade, Stark, & Allen, 1992; Jarodzka, Holmqvist, & Nyström, 2010), particularly when there is no clear hypothesis to guide more specific analyses. The algorithm presented in this chapter complements existing algorithms as an intuitive and simple way to obtain similarity ratings.

<div align="center">∞</div>

PART 1: VISUAL ATTENTION AND STABILITY

II. REVIEW: VISUAL ATTENTION AND STABILITY

Abstract — In the present review we address the relationship between attention and visual stability. Even though with each eye, head and body movement the retinal image changes dramatically, we perceive the world as stable and are able to perform visually guided actions. However, visual stability is not as complete as introspection would lead us to believe. We attend to only a few items at a time and stability is maintained only for those items. There appear to be two distinct mechanisms underlying visual stability. The first is a passive mechanism: The visual system assumes the world to be stable, unless there is a clear discrepancy between the pre- and post-saccadic image of the region surrounding the saccade target. This is related to the pre-saccadic shift of attention, which allows for an accurate preview of the saccade target. The second is an active mechanism: Information about attended objects is remapped within retinotopic maps to compensate for eye movements. The locus of attention itself, which is also characterised by localised retinotopic activity, is remapped as well. We conclude that visual attention is crucial in our perception of a stable world.

∞

In recent years many researchers have emphasised that vision is an active process (e.g., Findlay & Gilchrist, 2003; Merriam & Colby, 2005; O'Regan & Noë, 2001). This emphasis is well justified, since what we see depends as much on internal cognitive processes, as it does on what is actually out there to see. An important aspect of active vision is that of all the visual information that is available to us only a very limited selection is fully processed and ultimately guides action and perception. The remainder of the information is filtered out in the early stages of processing. This mechanism of selection is generally referred to as selective visual attention. By covertly attending (i.e., without making an eye movement) to a stimulus, we perceive that stimulus more clearly than we would if attention were unfocused or directed elsewhere. This increased perceptual ability can be measured as an increased sensitivity to faint stimuli (Bashinski & Bacharach, 1980), enhancement of perceived contrast (Carrasco, Penpeci-Talgar, & Eckstein, 2000), and decreased reaction times to attended stimuli (Posner, 1980). In addition, visual attention is characterised by an inhibitory surround: Processing of stimuli outside, but near

the focus of attention is suppressed (Bahcall & Kowler, 1999; Mathôt et al., 2010; Mounts, 2000a). These finding are paralleled by neurophysiological studies that have shown that visual attention enhances neural responsiveness and selectivity (Reynolds, Pasternak, & Desimone, 2000; Spitzer, Desimone, & Moran, 1988) and that the neural response to non-attended stimuli near the focus of attention is inhibited (Moran & Desimone, 1985; for a review, see Reynolds & Chelazzi, 2004). In addition to directing attention to a location in space, it is also possible to direct attention to objects (e.g., Roelfsema, Lamme, & Spekreijse, 1998; Theeuwes, Mathôt, & Kingstone, 2010) or based on non-spatial features, such as colour, direction of motion (Saenz, Buracas, & Boynton, 2002) and gestalt principles (Wannig, Stanisor, & Roelfsema, 2011). However, in the present review we will focus on spatial attention, which is especially relevant in the context of visual stability.

The effects of attention as studied in the laboratory are generally modest. For example, people respond about 20 ms faster to a validly cued, attended stimulus than to an uncued, neutral stimulus (Posner, 1980). Presumably, this effect is small, because the display is sparse. In such a display, there is little competition between stimuli and therefore little effect of attention (Reynolds et al., 2000). However, in more natural settings the effects of attention can be substantial. This has been elegantly demonstrated in experiments on change blindness (Rensink et al., 1997; Simons & Rensink, 2005; see also Mack, 2003; Mack & Rock, 1998). In a typical change blindness experiment, participants observe two displays that are presented in alternation and differ in some important respect. If the two pictures are presented in immediate succession the change is readily detected, because it constitutes a unique visual event. However, if a blank screen is introduced between the two displays, it takes considerable time and effort to detect the change. This is because the entire display now flashes and the change is no longer a unique visual event. In order to nevertheless find the changing element, you have to attend to different parts of the display in a serial fashion. This illustrates that, in natural settings, it is an understatement to say that attention provides us with improved perceptual abilities. Rather, we consciously perceive only what we attend to (Noë & O'Regan, 2000), which will be a recurring theme in the present review.

An equally important aspect of active vision is that we continuously make eye, head and body movements. This way we actively control the visual input that we receive, even prior to any effects of covert visual attention. Eye movements are an integral part of vision, because without eye movements we would only perceive a very small part of the visual field with high acuity and in colour: The part that projects onto the fovea. By making eye movements we sequentially extract information from different parts of the visual field.

This method of actively sampling our environment comes so natural that we are generally not aware of it. Perhaps even more surprisingly, we are also not aware of the fact that with each eye movement there is a corresponding shift in our retinal image of the world. Somehow, despite incomplete and unstable visual input, we feel as though we have a complete and stable percept of the world and, arguably even more importantly, we are able to effortlessly perform visually guided actions.

In the current review we focus on the role of attention in visual stability. In the first section we discuss trans-saccadic memory, a visual memory buffer that allows (some) information to be retained across saccades. The second section describes the assumption of stability: We perceive a stable world, simply because we assume the world to be stable. The final three sections discuss remapping of receptive fields, which has received considerable interest as a potential mechanism underlying visual stability. Sections three and four deal with neurophysiological and behavioural studies on remapping respectively. Section five describes a number of alternative views that challenge the traditional notion of remapping.

Trans-saccadic memory

Subjective experience suggests that visual stability is absolute and complete. Not surprisingly therefore, it has been suggested that conscious experience does not rely directly on retinotopically organised input, but on a representation of the world which is independent of eye position (spatiotopic). In general terms, trans-saccadic memory (TSM) is such a spatiotopic memory buffer. However, its exact characteristics have been the subject of substantial debate and revision (for a review, see Irwin, 1996). Initially, TSM was assumed to be a pre-attentive visual buffer, containing all visual detail of the world. In this form, it was also called an integrative visual buffer to emphasise its role in trans-saccadic integration (McConkie & Rayner, 1976; Trehub, 1977). Because trans-saccadic integration was believed to occur pre-attentively (at an early stage of processing) it was predicted that people should be able to seamlessly integrate information across saccades. Essentially, it should not matter whether people make eye movements or not. Although there was some initial support for this idea (Breitmeyer, Kropfl, & Julesz, 1982; Jonides, Irwin, & Yantis, 1982), further scrutiny revealed that people are often unable to integrate information across saccades (Bridgeman & Mayer, 1983; Irwin et al., 1983; O'Regan & Lévy-Schoen, 1983), whereas they have no difficulty doing so while fixating (Di Lollo, 1980). These findings did not cause the notion of TSM to be abandoned, but the concept clearly needed to be modified (see Figure ii.1).
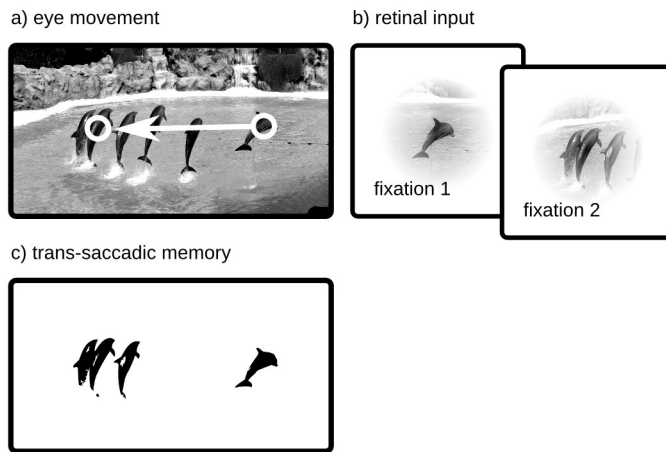
Figure ii.1. A schematic representation of trans-saccadic memory (TSM). a) A leftwards eye movement is executed. b) Visual input consists of two successive fixations. c) The two fixations are integrated in TSM. Since we generally do not attend to the background, no information about the background is retained (Prime et al., 2007). In addition, TSM contains mostly conceptual information (Irwin, 1996; but see e.g. Demeyer et al., 2009). For example, the fact that there are dolphins present in the scene is retained, but subtle differences in colouration are lost.

In a series of studies, Irwin (1991, 1992a, 1992b; Irwin & Andrews, 1996) investigated the properties of TSM. In one experiment participants were presented with an array of letters (Irwin, 1992b). Next a saccade target was presented. As soon as participants initiated an eye movement, the array of letters was extinguished. After the eye movement a cue was presented and participants reported which letter had been presented at the cued location. This experiment revealed two important properties of TSM. First, people remembered only 3 to 4 letters, suggesting a capacity limitation. In addition, memory was best for objects near the saccade target. The importance of this latter finding became apparent when later studies revealed that an eye movement is always preceded by a covert shift of attention (Deubel & Schneider, 1996; Godijn & Pratt, 2002; Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995), so that the saccade target receives an attentional benefit. This explained why in Irwin's (1992b) study TSM was best for stimuli near the saccade target: Those stimuli received an attentional benefit and were therefore stored in TSM. The idea that attention functions as a 'gatekeeper' for TSM was investigated in more detail by Prime and colleagues (Prime, Tsotsos, Keith, & Crawford, 2007; see also Melcher, 2008b). They instructed participants to remember a

number of randomly positioned stimuli (patches of tilted lines known as Gabor patches). One of these stimuli was cued prior to its presentation, indicating that it was likely to be probed in the response phase. Presumably, participants attended to the cued stimulus. After an eye movement a probe stimulus was presented (another Gabor patch). Participants reported whether the probe was tilted clockwise or counter-clockwise, relative to the original stimulus (the stimulus which had previously been presented at the same location). The crucial finding was that performance was best for stimuli which had been cued, confirming that TSM is best for attended stimuli.

On the basis of these findings it can be concluded that TSM has a limited capacity and that attention acts as a 'gatekeeper'. Other properties, not directly related to visual attention, are that TSM deals predominantly with abstract, conceptual information (Henderson, Pollatsek, & Rayner, 1987; Irwin, 1996) and has a coarse spatial resolution (Pollatsek, Rayner, & Henderson, 1990). Low-level, non-conceptual information has some effect on trans-saccadic integration, the extent of which is a matter of debate (e.g., Demeyer, De Graef, Wagemans, & Verfaillie, 2009, 2010), but there appears to be a type of gradient: Low-level features are not entirely lost, but conceptual features are dominant (Melcher, 2005). Taken together, the properties of TSM are strongly reminiscent of spatial working memory. The natural conclusion is that TSM is not a separate entity, but simply a name for spatial working memory in the context of eye movements (Irwin, 1991).

To conclude, researchers have posited the existence of trans-saccadic memory (TSM). TSM contains a spatiotopic representation of the world, which is independent of eye position. In order to be integrated across saccades, stimuli need to be stored in TSM. Rather than a dedicated mechanism for trans-saccadic perception, TSM appears to rely on working memory. TSM has a limited capacity and only information about attended stimuli is retained (Irwin, 1991; Melcher, 2008b; Prime et al., 2007).

The assumption of stability

As was mentioned in the previous section, every saccade is preceded by a covert shift of attention (Deubel & Schneider, 1996; Godijn & Pratt, 2002; Hoffman & Subramaniam, 1995; Kowler et al., 1995). In a typical paradigm investigating pre-saccadic shifts of attention, participants are instructed to make an eye movement to a particular location. After participants have been cued to make a saccade, but before the eyes set in motion, a stimulus is presented at the saccade goal. The pre-saccadic shift of attention is reflected by the finding that stimuli presented at the saccade goal are more readily discriminated (Deubel & Schneider, 1996) and elicit stronger priming effects (Godijn & Pratt, 2002) than stimuli presented elsewhere. A related finding is that people subjectively feel that the eyes

have already moved to the saccade target, when in fact the saccade is yet to be executed (Deubel, Irwin, & Schneider, 1999; Hunt & Cavanagh, 2009). Presumably, this is due to the pre-saccadic shift of attention, which provides improved perception of the saccade target before it has been foveated.

A number of researchers have suggested that the pre-saccadic shift of attention is integral to visual stability (Currie et al., 2000; Deubel, Bridgeman, & Schneider, 1998; Deubel & Schneider, 1994; McConkie & Currie, 1996). In this view, attention precedes an eye movement to allow for an accurate preview of the saccade target. After the eye movement, this region is observed again and trans-saccadic integration occurs based on the assumption that the saccade target and its surroundings have remained stable. It is, in a sense, a snapshot theory, in which pre- and post-saccadic snapshots are superimposed. This differs from the traditional notion of trans-saccadic memory in that no knowledge of absolute spatial positions is required, since snapshots are integrated based on content and local structure rather than absolute location. This also differs from the integrative visual buffer in that these snapshots are believed to contain mostly abstract representations, modulated by attention.

Assuming that the saccade target is stable (at least for the duration of a saccade) makes ecological sense, but in the lab it can be violated quite easily by moving the saccade target while the eyes are in motion. Since visual perception is strongly suppressed during eye movements (Holt, 1903; Matin, 1974; for a recent review, see Castet, 2010), the exact moment of displacement is not observed and the visual system relies on pre- and post-saccadic snapshots to detect the displacement. Remarkably large displacements of the saccade target go unnoticed (Bridgeman et al., 1975), confirming the notion that the visual system assumes the saccade target to be stable unless there is strong evidence to the contrary. In situations where the saccade target is clearly not stable, for example if the saccade target is already in motion prior to the saccade (Gysen, De Graef, & Verfaillie, 2002) or is briefly blanked after the saccade (Deubel et al., 1996), displacement detection is greatly improved.

Visual attention is intricately related to the assumption of stability, as attention appears to be a determining factor in which objects are assumed to be stable. We can illustrate this by describing the assumption of stability in terms of 'finding the best fit' (see Figure ii.2). As mentioned, pre- and post-saccadic snapshots of the saccade goal and its surroundings are constructed. These snapshots contain representations of stimuli to the extent that they are attended (an "attentional landscape", cf. Baldauf & Deubel, 2010). Effectively, this means that the saccade target itself is strongly represented, but nearby stimuli can also be
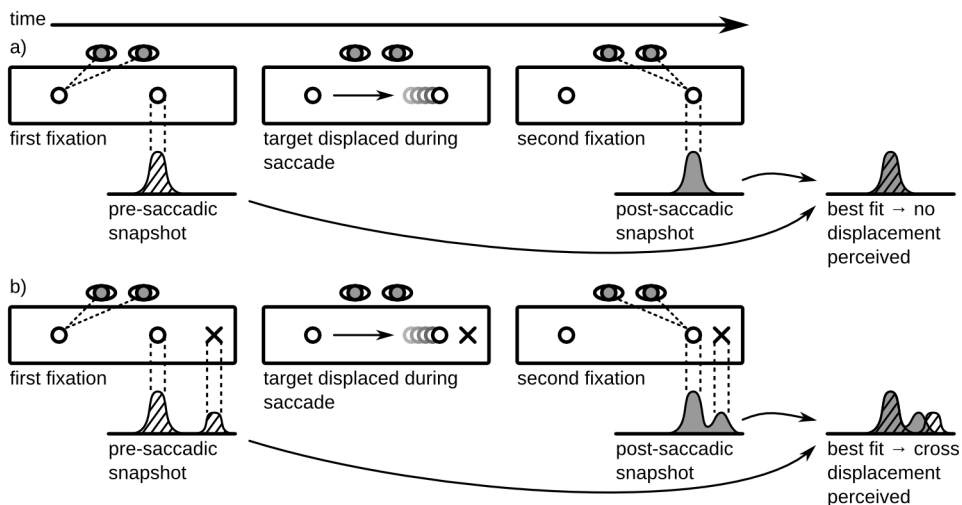
Figure ii.2. A description of the assumption of stability in terms of 'finding the best fit'. a) The saccade target is displaced during the saccade. Because there is nevertheless a perfect fit between the pre- and post-saccadic snapshots, the displacement is not perceived (Bridgeman et al., 1975). b) The saccade target is displaced, while an additional stimulus (an X) remains stable. The saccade target is more strongly attended than the X and therefore has a larger 'bump' of activation. Consequently, the best fit means matching the pre- and post-saccadic saccade target, causing a mismatch of the pre- and post-saccadic X. As a result, the X is erroneously perceived as being displaced (Heywood & Churcher, 1981).

represented, although more weakly. Integration occurs based on the assumption that the best fit between the pre- and post-saccadic snapshots is the true fit. This simple principle explains many findings. For example, if the saccade target is displaced during the saccade, there is still a perfect fit between pre- and post-saccadic snapshots (see Figure ii.2a). The only difference lies in absolute spatial position, which is not a factor in determining the best fit. Consequently, the visual system fails to detect the displacement. We can also consider what happens if a second stimulus (an X) is added, which remains stable while the saccade target is displaced (see Figure ii.2b). In this case, the best fit still results from matching the pre- and post-saccadic saccade target. The best fit requires a misalignment of the pre- and post-saccadic X, because it receives less attention than the saccade target and therefore contributes less to the overall fit. Consequently, the cross is erroneously perceived as being displaced (Heywood & Churcher, 1981). This principle also explains why, if multiple stimuli are presented, a displacement is generally attributed to the stimulus which is briefly blanked at the moment the eyes arrive at the saccade target, regardless of which stimulus was actually displaced (Deubel et al., 1998; Deubel, Koch, & Bridgeman,

2010). This is because only the stimuli that are present right after the saccade contribute to the fit. If one of the stimuli is missing (because it has been blanked) the fit will be poor, but the best fit will nevertheless result from aligning the stimuli that are present.

There are a number of qualifications that should be made. First, we have not considered what happened if a stimulus is replaced by a qualitatively different stimulus during a saccade. Changing stimulus identity has a definite effect on trans-saccadic integration, which indicates that qualitative factors are important in matching pre- and post-saccadic information (e.g., Henderson et al., 1987). In addition, even if a stimulus is briefly blanked after the saccade it may still serve as a stable reference point, provided that other stimuli are blanked for a longer period of time (Deubel et al., 2010). This suggests that there is substantial temporal fuzziness in the assumption of stability. Perhaps even more surprisingly, effects of stimulus blanking and displacement can also be observed during fixation, suggesting that the assumption of stability is a general phenomenon and not strictly limited to trans-saccadic perception (Deubel et al., 2010).

An important question is: If only a saccade target is presented, why does post-saccadic blanking improve detection of its displacement (Deubel et al., 1996)? The fact that blanking breaks the assumption of stability is part of the explanation, but leaves us with another question: Why do we still have a sense of position when we cannot rely on the assumption of stability? The answer must be that we fall back on different mechanisms (see the sections on *Remapping and Attention*). This is also supported by evidence from corrective saccades. If a saccade target is displaced during the saccade, corrective saccades are executed towards the new location of the saccade target (Hollingworth, Richard, & Luck, 2008). This is the assumption of stability at work. However, if the saccade target is removed (after the eyes set in motion), corrective saccades are executed towards the former location of the saccade target (Shebilske, 1976). Clearly, the visual system has a way of maintaining positional informations across saccades that does not rely on the assumption of stability.

To conclude, our visual system exploits the fact that the world is a stable place, at least for the duration of an eye movement (Deubel & Schneider, 1994; McConkie & Currie, 1996; also see Chapter vii of this thesis). Generally, the saccade target dominates the assumption of stability, because it is strongly attended just before each eye movement (e.g., Kowler et al., 1995), but other attended stimuli may serve as stable reference points as well.

Remapping and attention: Neurophysiology

As visual information enters the primary visual cortex, retinal topography is preserved: Adjacent neurons process information from adjacent, and usually largely overlapping parts of the retina (Talbot & Marshall, 1941). However, as we move further upstream in the visual processing hierarchy, things become considerably less clear. Receptive fields (RFs) of neurons in these later areas differ in many important respects, but here we focus on the distinction between retinotopy and spatiotopy. In addition, RFs change in different ways in the interval preceding an eye movement (Tolias et al., 2001), but here we restrict the discussion to pre-saccadic RF-shifts in the direction of the eye movement, usually called predictive remapping.

If the RF of a neuron is retinotopic it is anchored to a location on the retina, which may correspond to different locations in the world depending on eye position. This is essentially what underlies the problem of visual stability. In contrast, if a neuron has a spatiotopic RF it is always responsive to the same spatial location, irrespective of eye, body, and head position. Because in most studies the head and body are in a fixed position, the term 'spatiotopic' is often used loosely and applied to responses that are highly independent of eye position. An important question is whether spatiotopy exists in the brain. It is attractive to assume that it does, since this would effectively solve the problem of visual stability. According to the spatiotopic hypothesis, action and conscious experience are based on spatiotopically organised brain areas. This bears some conceptual resemblance to trans-saccadic memory, although trans-saccadic memory is a cognitive construct which is not necessarily intended to reflect a spatiotopic map at the neural level.

In apparent support of the spatiotopic hypothesis, brain areas have been identified in which RFs are modulated by eye position (Crespi et al., 2011; d' Avossa et al., 2007; Duhamel, Bremmer, BenHamed, & Graf, 1997; Galletti, Battaglini, & Fattori, 1995; but see Gardner et al., 2008). RFs in these areas are not retinotopic, but neither is it obvious that they are of the fine grained spatiotopic sort that would be expected based on the spatiotopic hypothesis. An alternative, perhaps more plausible interpretation is that these RFs are tailored towards a specific modality, rather than being spatiotopic and directly related to visual stability. For example, in the extended dorsal stream there is a continuum from visual to motor responses, such that observing an object automatically activates an associated motor program (Fadiga, Fogassi, Gallese, & Rizzolatti, 2000). Since information in retinal coordinates is of little use for programming manual reaching movements, a translation from retinotopic coordinates to a more appropriate frame of reference (for example, body-centred coordinates) seems natural (see e.g., Overvliet,

Azañon, & Soto-Faraco, 2011). However, this does not require true spatiotopy and does not provide strong evidence for the spatiotopic hypothesis.

For this reason, the spatiotopic hypothesis has fallen out of favour as the complete solution to the problem of visual stability (see Wurtz, 2008 for a discussion). However, it is well established that many RFs are modulated by eye position, presumably mediated by a corollary discharge (Sommer & Wurtz, 2008). It has been proposed that remapping of RFs might be the solution to the problem of visual stability. Before discussing neurophysiological studies we will briefly introduce the concept of remapping by analogy.

Imagine that you are sitting in a train without windows. You are instructed to remain at the same position—not relative to the train, but relative to the outside world. This is tricky, because the train occasionally moves and you cannot look out of the windows to see where you are. Fortunately, the train operator always announces exactly how far and in what direction the train is going to move, just before the train actually sets in motion. Therefore, if you hear "Folks, we are about to move 20 meters forward", you quickly run 20 meters to the back of the train, thus compensating for the movement of the train.

How does this example relate to visual stability? Imagine that a stimulus is briefly presented. Even after the stimulus has been extinguished, there is some residual neural activity. This is often called a memory trace (cf. Merriam, Genovese, & Colby, 2003), but you can also think of it as an attention-related increase in baseline activity (Colby, Duhamel, & Goldberg, 1996). The problem that the memory trace faces is analogous to that of our example. If the eyes move, the memory trace becomes misaligned with the world: The same spot in the retinotopic map now corresponds to a different location in the real world and therefore the memory trace is not sitting in the right spot of the retinotopic map any more. Fortunately, the corollary discharge informs the visual system of the impending eye movement. Using this information, the memory trace can be transferred onto a different set of neurons in the same retinotopic map, so that it remains correctly aligned with the world (Merriam et al., 2003). This mechanism is called remapping or spatial updating. In a nutshell, remapping is a transfer of activity between retinotopically organised neurons. This transfer of activity is such that it compensates for eye movements, effectively updating retinotopic representations to prevent a misalignment with the world. This provides a way for the visual system to maintain visual stability without the need for spatiotopic RFs, and therefore it is sometimes called the retinotopic hypothesis.

Remember that the train operator signals movement before the train actually sets in motion. This allows you to get a head start, by running to the back of the train before the train starts moving forward. Similarly, a corollary discharge informs the visual system of
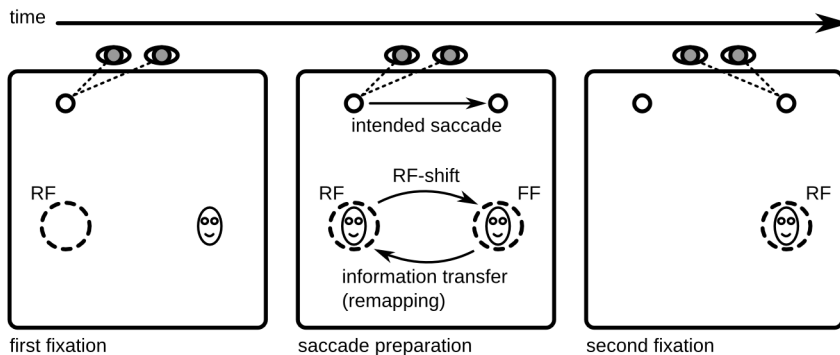
Figure ii.3. A schematic representation of predictive remapping (Duhamel et al., 1992). a) During fixation a given neuron is responsive to a single part of the visual field, its receptive field (RF). b) As a saccade is being prepared, but before it has been executed, the neuron also becomes responsive to the location that will be brought into its RF, its future field (FF). Effectively, the neuron takes a 'sneak peak' at its FF, which allows it to take over whatever activation is there. This activation may represent a physical stimulus (such as the face presented here), but also attentional activation (Mathôt & Theeuwes, 2010a) or a memory trace of a stimulus which is no longer present (Merriam et al., 2003). The activation in the FF is therefore being transferred (remapped) in the direction opposite from the RF shift. c) After the eye movement the FF has become the RF of the neuron.

an eye movement before it occurs, since it conveys information about intended rather than actual eye movements. This allows remapping to start before an eye movement, in which case it is referred to as predictive remapping. So far we have looked at remapping from the perspective of the memory trace (of course the same principles apply to remapping of visual information in general). However, predictive remapping is commonly described in terms of RFs. This distinction is important, because the identity of a memory trace is independent of the neurons that encode it. After all, the memory trace may be remapped from one set of neurons onto another. This shift in perspective is also useful, because it sheds some light on how remapping works. In the interval preceding an eye movement, RFs shift in the direction of the eye movement (Duhamel et al., 1992). This may seem at odds with the fact that the memory trace is remapped in the direction opposite from the eye movement (as you run against the movement of the train), but it is not (see Figure ii.3). The anticipatory RF-shift allows a neuron to take a 'sneak peak' at the location that will be brought into its RF. This is somewhat analogous to the pre-saccadic shift of attention (see the previous section) but applies to the visual field as a whole, rather than just the currently fixated location. In this context, the RF-location-to-be is often called the future field (FF). If the memory trace happens to be in a neuron's FF, the neuron will take over some of the

memory trace activity, which corresponds to remapping of the memory trace. Remapping of activity is therefore in the direction opposite from the anticipatory RF-shift.

We now move on to the actual neurophysiological studies. The first evidence for remapping was reported in primate single-cell recording studies of the frontal eye fields (FEF; Goldberg & Bruce, 1990) and the lateral intraparietal area (LIP; Barash, Bracewell, Fogassi, Gnadt, & Andersen, 1991), using the double-step paradigm. In a typical double-step task, two saccade targets are briefly presented. After the targets have been removed, participants (monkeys in this case) make two successive eye movements to where the targets used to be. The rationale behind this paradigm is that the first eye movement causes a retinal displacement of the location of the second target. Because the second target is no longer visible at the time of the second eye movement, somehow this retinal displacement needs to be taken into account when programming the second eye movement. The crucial finding was that if the location of the second target was brought into a neuron's RF (or movement field) by the first saccade, the neuron would often respond, even though the second target was no longer visible. The explanation is that the memory trace of the second target was remapped to compensate for the eye movement, and that the neuron was responding to the remapped memory trace.

In a landmark study, Duhamel, Colby and Goldberg (1992) extended these finding in a remarkable way. They recorded cells from the monkey LIP. The crucial finding was that almost half the neurons became responsive to their FF after the monkey had been instructed to make a saccade, but before the saccade had been executed: unmistakable evidence for predictive remapping. In addition, neurons became less responsive to their current RF: RFs shifted from current to future field. However, later studies showed that in other areas neurons sometimes become responsive to their FF, but remain responsive to their current RF as well (Nakamura & Colby, 2002). When a stimulus was removed before the saccade, Duhamel and colleagues (1992) found evidence for remapping (not necessarily predictive) of the memory trace of the removed stimulus. This was the case for almost all LIP neurons, so remapping is really a ubiquitous phenomenon in some brain areas.

In addition to the parietal cortex (Duhamel et al., 1992), remapping has been demonstrated in the FEF (Sommer & Wurtz, 2006), the superior colliculus (Walker, Fitzgibbon, & Goldberg, 1995) areas V3, V2 (Nakamura & Colby, 2002) and even in V1 (Khayat, Spekreijse, & Roelfsema, 2004). Despite the fact that remapping occurs at many, if not all levels of the visual system, the tendency is for early visual areas to show less remapping and later in time than areas such as the FEF (Merriam, Genovese, & Colby,

2007). This, and the observation that the FEF receive a strong corollary discharge (Sommer & Wurtz, 2004), has led researchers to suggest that the FEF may be an important source for remapping of visual responses (Sommer & Wurtz, 2003).

With respect to visual attention an important question is whether a covert shift of attention is by itself sufficient to trigger remapping, as one might think given the strong link between visual attention and the oculomotor system (Craighero & Rizzolatti, 2005). This is not the case: Remapping occurs only in combination with eye movements and there is no evidence to suggest that it can be induced by a covert shift of attention (Colby, 1996). This makes sense, of course, because remapping in the absence of an eye movement would cause retinotopic maps to become misaligned with the world, which would stand in contrast with the assumption that remapping is a mechanism to prevent misalignment. However, visual attention does play an important role in remapping in a different way. By recording neurons from LIP, Gottlieb, Kusunoki and Goldberg (1998) investigated how remapping is affected by attention. Area LIP is often conceptualized as a priority (or saliency) map (Bisley & Goldberg, 2010). That is, it is believed to contain little information about specific features, such as colour and form, but to be driven by the abstract notion of 'priority'. The priority of an object is determined by bottom-up and top-down factors. Bottom-up factors are due to stimulus features, such as a sudden onset or a conspicuous colour, but are short-lasting (Donk & van Zoest, 2008; Siebold, van Zoest, & Donk, 2011): An onset stimulus initially captures attention, but attention can be disengaged quickly. Top-down factors are due to the behavioural relevance of an object and can be long-lasting: If you want to, you can attend for a long time to a stimulus, even if it is inconspicuous. The priority map is 'read out' by the visual system to guide attention and therefore there is a strong correspondence between activation in the priority map and the allocation of attention (Bisley & Goldberg, 2010). In accord with the view of LIP as a priority map, Gottlieb and colleagues (1998) found that LIP neurons showed sustained response to a behaviourally relevant stimulus (a saccade target), brief response to a behaviourally irrelevant onset stimulus, and little to no response to a behaviourally irrelevant persistent stimulus (i.e., a 'boring' stimulus that has been visible for an extended period of time). With respect to visual stability, an important question is what happens if a persistent stimulus is brought into a neuron's RF. From the neuron's perspective, which has never 'seen' the persistent stimulus before, the stimulus is novel and might therefore elicit a burst of activity as though it were an onset. This would result in a large number of pseudo-onset stimuli with every saccade, which would clearly be detrimental to performance and our sense of visual stability. What Gottlieb and colleagues (1998) found

was that a stimulus elicits a burst of activity only once, even if an eye movement brings it into the RFs of a new population of neurons. This shows that an important characteristic of bottom-up attention is preserved across saccades: Stimuli capture attention only once.

Another important question is whether remapping applies to all stimuli or only to a subset. The study by Gottlieb and colleagues (1998) shows that stimuli which are attended are also remapped. Of course it is difficult to show conclusively that information about unattended objects is never remapped, but many researchers believe this to be the case. Therefore, since most stimuli are not attended, most stimuli are not remapped. This is strongly reminiscent of the behavioural studies that we discussed previously, showing that trans-saccadic memory is best for attended stimuli (Irwin & Gordon, 1998; Melcher, 2008b).

To conclude, remapping (or spatial updating) is a strong candidate mechanism for visual stability (Wurtz, 2008). Remapping refers to the transfer of visual information within retinotopic maps to compensate for eye movements (Duhamel et al., 1992). It is generally believed that remapping is limited to attended stimuli (Gottlieb et al., 1998). Therefore, visual stability is maintained only for those stimuli that guide action and conscious perception.

Remapping and attention: Behavioural findings

The hypothesis that remapping of receptive fields (RFs) underlies visual stability is originally based on neurophysiological findings. However, there is a fast growing body of behavioural research on remapping. In this section we highlight a number of behavioural studies which have specifically investigated the role of visual attention in remapping.

Melcher (2008b) investigated trans-saccadic integration using the tilt-adaptation after-effect (TAE). In a typical TAE experiment, participants are exposed to a tilted grating (the adapter) for some time. Next they are presented with another, slightly tilted grating (the tester) and are asked to report the orientation of the tester. TAE is a bias to report the tester as being tilted away from the adapter orientation. TAE persists, albeit in slightly reduced form, if an eye movement is executed between the presentation of the adapter and the tester, if they are presented at the same spatial location (Melcher, 2005, 2007, 2008a, 2008b; however, see Knapen, Rolfs, Wexler, & Cavanagh, 2010 and Chapter vi of this thesis). This suggests that the representation of the adapter is remapped to compensate for the eye movement (see also Cha & Chong, 2010). Gottlieb and colleagues (1998) have shown that, at least for LIP neurons, visual attention determines which objects are represented and consequently remapped. Similarly, Melcher (2008b) found that if attention was directed to an adapter stimulus, TAE increased. However, this was the case regardless

of whether an eye movement had been made between the presentation of the adapter and the tester. Again, this demonstrates that attention determines which objects are represented and that only represented (i.e., attended) objects are remapped. The role of attention in visual stability is therefore the same as the more general role of attention as a perceptual filter.

Another important question is whether attention itself is remapped. Even though the locus of attention is not a physical stimulus, it is characterised by localised activity in the visual system and as such can be remapped like a regular stimulus. It has generally not been described in these terms, but this is exactly what has been done in the previously discussed neurophysiological and neuroimaging studies, which investigated remapping of a memory trace (e.g., Duhamel et al., 1992; Goldberg & Bruce, 1990; Merriam et al., 2003). In these studies a stimulus was presented briefly, presumably attracting attention. Even after the stimulus had been removed some residual activity was observed. This residual activity is usually referred to as a memory trace, but as suggested earlier it can also be thought of as an attention related increase in activity (Colby et al., 1996).

Posner and Cohen (1984) were the first to investigate the reference frame of attention or, using modern terminology, remapping of attention. They investigated both attentional facilitation and the subsequent inhibitory phase (Inhibition of Return, IOR). Posner and Cohen (1984) found that facilitation was retinotopic: If participants made an eye movement, the locus of attention moved with the eyes to a new spatial position. In contrast, they found that IOR was spatiotopic: The locus of inhibition remained at the same spatial location regardless of eye movements (see also Maylor & Hockey, 1985; but see Abrams & Pratt, 2000). The finding that IOR is spatiotopic makes ecological sense, because IOR is a relatively sustained effect, typically spanning multiple eye movements (Klein & MacInnes, 1999; but see Ludwig, Farrell, Ellis, & Gilchrist, 2009). However, the dissociation between facilitation (retinotopic) and inhibition (spatiotopic) was surprising, since these two phenomena are generally assumed to be linked.

More recently, Golomb and colleagues (2008) investigated remapping of attention in more detail. In order to attract attention to a location, they instructed participants to remember the location of a briefly flashed cue (cf. Awh, Jonides, & Reuter-Lorenz, 1998). After participants had made an eye movement, a line-segment was presented at one of three locations: the original attended location, a location that retinotopically matched the original attended location, or a control location. The reaction time difference between the location of interest and the control location in reporting the orientation of the line-segment (attentional facilitation) was taken as a measure of attentional allocation. The results

depended strongly on the task instruction and on the moment at which the line-segment was presented. If the instruction was simply to remember the cued location, facilitation was initially strongest at the retinotopic location. However, retinotopic facilitation dissipated quickly, whereas spatiotopic facilitation was more sustained. This suggests that the locus of attention was remapped to compensate for the eye movement, resulting in spatiotopic facilitation. Because remapping was incomplete, there was retinotopic facilitation directly after the eye movement, which dissipated rapidly due to a lack of maintenance. However, if the instruction was to remember the location relative to the eyes, the results were quite different. In this case, there was sustained retinotopic facilitation and even a hint of spatiotopic inhibition. This led the authors to conclude that the locus of attention is essentially tied to retinotopic coordinates, and is not remapped unless this is explicitly required. In other words, the authors propose an additional restriction on remapping: Even attended objects are remapped only when this is required for the task at hand. However, to memorise a location relative to the eyes is arguably an awkward instruction and participants may have resorted to unknown strategies in order to comply. The instruction to simply memorise a location is more natural and indeed yielded results which are more consistent with neurophysiological studies, which typically do not consider task-instruction at all (e.g., Duhamel et al., 1992).

In a paradigm inspired by the study by Golomb and colleagues (2008), we likewise investigated the effect of an eye movement on the locus of attention (Mathôt & Theeuwes, 2010a). We presented an onset stimulus, which is known to attract attention (Mathôt & Theeuwes, 2012; Yantis & Jonides, 1984). In one experiment, we presented the probe (which was also a line-segment) at the time of the saccade, in which case we found facilitation at both the spatiotopic and retinotopic location. Again, this suggests that the locus of attention is remapped, resulting in spatiotopic facilitation, but that remapping is only partial, resulting in retinotopic facilitation. In a related experiment (Mathôt & Theeuwes, 2010b; van Koningsbruggen, Gabay, Sapir, Henik, & Rafal, 2010; but see Pertzov, Zohary, & Avidan, 2010), participants did not respond manually to a probe, but, after the first eye movement, made a second eye movement to the location of interest (spatiotopic, retinotopic or one of two locations). The first eye movement allowed us to dissociate retinotopic and spatiotopic coordinates. The latency of the second eye movement was used as a measure of attention: Faster eye movements indicate more attention (e.g., Kowler et al., 1995). Because of the relatively long interval between onset presentation and the second saccade, we expected IOR rather than facilitation. The results were clear-cut: If the second saccade was made right after the first saccade, IOR was

predominantly retinotopic. At longer intervals IOR was predominantly spatiotopic. These findings resemble those of Golomb and colleagues (2008), who reported the same pattern of results for attentional facilitation. In relation to the studies of Posner and Cohen (1984), these findings illustrate that reference frames are flexible and dynamic: Effects may appear to be retinotopic or spatiotopic, depending on when you probe. This may account for the apparent dissociation between attentional facilitation (retinotopic) and IOR (spatiotopic).

In another experiment (Mathôt & Theeuwes, 2010a) we presented the probe stimulus after observers were instructed to make a saccade, but before the saccade had been executed (see Figure iii.3). The rationale behind this experiment was as follows: We assumed that the presentation of the onset stimulus excited a population of neurons (Gottlieb et al., 1998). If a probe is subsequently presented within the receptive fields (RFs) of these excited neurons, processing of the probe is facilitated. Under normal circumstances, this means facilitation for probes presented at the same location as the onset. However, in the pre-saccadic interval a proportion of neurons become transiently responsive to their future field (FF; Duhamel et al., 1992). If a probe would be presented within the FFs of the neurons that were excited by the onset, facilitation should, in theory, be observed. Therefore, in some trials we presented the probe at the 'future-retinotopic' location that fell within these presumed FFs. Crucially, we found attentional facilitation for probes presented just before the saccade at the future-retinotopic location. This suggests that predictive remapping affects the locus of attention in the interval preceding saccade execution[6].

Another important question is whether an eye movement causes an attentional 'spread' or 'split'. A recent study shows that the spatiotopic and retinotopic loci of attention form two non-contiguous locations, suggestive of a split (Golomb, Marino, Chun, & Mazer, 2011), which is exactly what you would expect based neurophysiological evidence (Sommer & Wurtz, 2006).

In summary, behavioural findings on remapping and attention are consistent with neurophysiological evidence. There are two important conclusions. First, attention determines which stimuli are remapped (Melcher, 2008b). This is an efficient strategy,

---

6   Shortly after the publication of this review in the *Philosophical Transactions of the Royal Society B* a paper by Rolfs and colleagues (2011) appeared that criticised our study and others (Biber & Ilg, 2011; W. Harrison, Remington, & Mattingley, 2012; Melcher, 2007) for probing the incorrect location. Following the logic outlined in the previous section (and see Figure ii.3) they argued that one should expect attention to shift in the direction opposite from the saccadic eye movement. From an information processing perspective they are, of course, correct. But the logic outlined in this section illustrates that one can predict a shift in either direction. This is puzzling and presumably reflects an incomplete understanding of how predictive remapping works (Hamker, Zirnsak, Calow, & Lappe, 2008; Morris, Kubischik, Hoffmann, Krekelberg, & Bremmer, 2012).

because it limits the problem of visual stability to those objects for which it is truly a problem: attended objects, which we act upon and consciously perceive. Second, the locus of attention itself is remapped similar to remapping of a physical stimulus (Mathôt & Theeuwes, 2010a). Remapping is not an instantaneous process and a gradual shift from retinotopic to spatiotopic coordinates can be observed (Golomb et al., 2008; Mathôt & Theeuwes, 2010b).

Remapping and attention: Alternative interpretations

Not all researchers agree that the findings discussed in the previous sections should be interpreted as evidence for remapping of receptive fields (RFs). Here we discuss two divergent interpretations of the available data, which invoke the concept of attention in different ways.[7]

Hamker and colleagues have constructed a computational model of peri-saccadic RF-changes (Hamker, Zirnsak, Calow, & Lappe, 2008). By simulating single-cell recording studies they have shown that their model produces output consistent with empirical data (Zirnsak, Gerhards, Kiani, Lappe, & Hamker, 2011; Zirnsak, Lappe, & Hamker, 2010). Importantly, their model does so without incorporating predictive remapping in the sense that cells become selectively responsive to their future field (FF). Rather, their model relies on RF-shifts towards the saccade target (Tolias et al., 2001). For selective parts of the visual field this results in RF-shifts which resemble predictive remapping, but this is an illusion (see Figure ii.4). Because of these shifts, the number of RFs that encompass the saccade target increases, which results in increased capacity for processing the saccade target. This could correspond to the pre-saccadic shift of attention. Essentially, in this model all peri-saccadic RF-changes are ultimately linked to the pre-saccadic shift of attention. This compelling model explains many findings in a parsimonious way, although it fails to account for some results as well. Notably, the finding that FF and RF are non-contiguous areas is not easily explained (Sommer & Wurtz, 2006) as is the finding that, depending on stimulus configuration, the locus of attention may predictively shift to a location beyond the saccade target (Mathôt & Theeuwes, 2010a).

Cavanagh and colleagues (2010) propose yet another view on remapping. According to them remapping is best explained as predictive shifts of attention. They argue that just before a stimulus is brought into a neuron's RF, the neuron becomes more active in order to

---

7   Another alternative that has emerged very recently is based on neural coding of eye position (Morris et al., 2012). This model, which I think is very compelling, is discussed in the General discussion of Chapter v.
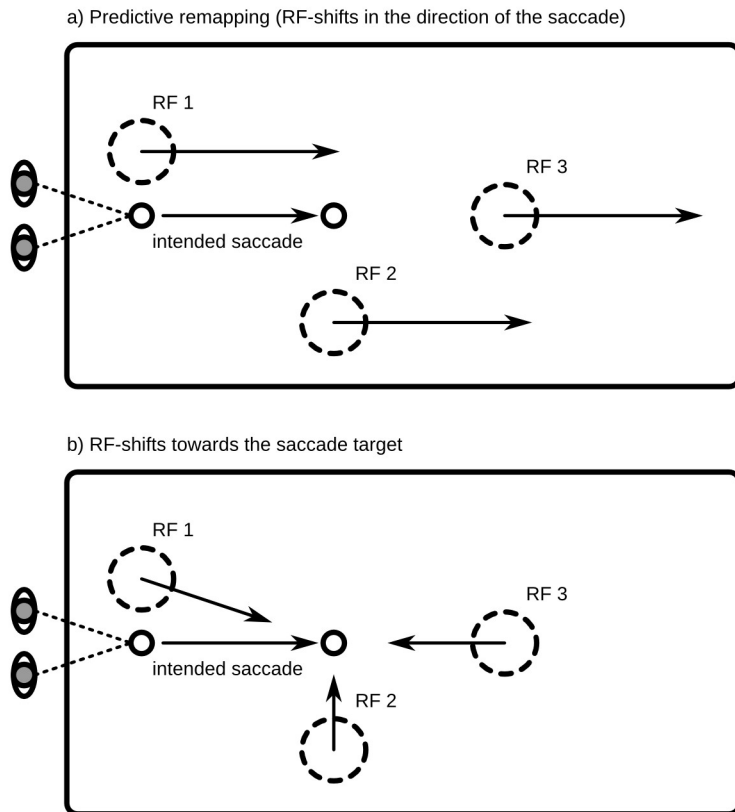
a) Predictive remapping (RF-shifts in the direction of the saccade)

RF 1

RF 3

intended saccade

RF 2

b) RF-shifts towards the saccade target

RF 1

RF 3

intended saccade

RF 2

Figure ii.4. A comparison between predictive remapping (a) and receptive field (RF)-shifts towards the saccade target (b). The model by Hamker and colleagues (Hamker et al., 2008; Zirnsak et al., 2011) shows that many neurophysiological findings which have been interpreted as predictive remapping can be explained in terms of RF-shifts towards the saccade target. You can see why this is the case by looking at RF 1. Careful probing of many locations is required in order to distinguish whether RF 1 shifts according to (a) or (b). For RFs 2 and 3 the distinction between (a) and (b) is much clearer.

prepare for the incoming information. Traditionally, information is believed to be transferred within retinotopic maps so that we do not need to re-acquire visual information after every saccade. According to Cavanagh and colleagues (2010) information is not preserved across saccades, but attentional shifts facilitate the process of re-acquiring what has been lost. Based on neurophysiological studies this is difficult to prove or refute, since remapping is typically investigated without taking stimulus features into account. However, if they are correct there should be no spatiotopic after-effects, since that would

indicate remapping of stimulus features. As pointed out by Cavanagh and colleagues (2010) a number of studies have indeed failed to show spatiotopic after-effects (Afraz & Cavanagh, 2009; Knapen et al., 2009, 2010; Wenderoth & Wiese, 2008; also see Chapter vi of this thesis). However, there are also several studies that have shown clear spatiotopic after-effects (Biber & Ilg, 2011; Ezzati et al., 2008; Melcher, 2005, 2008a, 2008b; Wittenberg et al., 2008) and trans-saccadic integration of object features (Demeyer et al., 2009, 2010; Melcher & Fracasso, 2012).

Conclusion

Over the years, research on visual stability has made considerable progress and a number of conclusions can be drawn. First, visual stability is not as absolute as introspection would lead us to believe. Stability is preserved only for a limited number of attended objects (Gottlieb et al., 1998; Irwin, 1991; Melcher, 2008b; Prime et al., 2007), which is sufficient since those objects guide action and conscious perception (Rensink et al., 1997). The feeling that we have a complete and stable perception of the entire visual field has been called a 'grand illusion' (Noë & O'Regan, 2000)[8].

Second, remapping appears to be one of the underlying mechanisms in visual stability (but see Cavanagh et al., 2010; Hamker et al., 2008). To compensate for eye movements, visual information is remapped within retinotopic maps. Although not all visual information is encoded retinotopically (Crespi et al., 2011; Duhamel et al., 1997; Galletti et al., 1995), there is little evidence to suggest, and really no reason to a priori assume, that true spatiotopy exists (Wurtz, 2008). The characteristics of trans-saccadic memory (TSM) as revealed by behavioural experiments strongly resemble the characteristics of remapping. As mentioned earlier, attention-gated limited capacity is a feature of both TSM and remapping (Gottlieb et al., 1998; Melcher, 2008b). In addition, the fact that TSM contains mostly, although not exclusively, conceptual information (Irwin, 1991; Melcher, 2005) is compatible with the finding that remapping occurs predominantly in higher visual areas, and is much less pronounced in visual areas dealing with those low-level features which are not readily integrated across saccades (Merriam et al., 2007).

Third, attention is not only involved in visual stability in a supervisory manner, but is itself the subject of remapping. In the pre-saccadic interval, the focus of attention is remapped predictively (Mathôt & Theeuwes, 2010a). Remapping of attention continues

---

8   In fact, as I understand it now, their argument is a bit more intricate. As described in the introduction,
     O'Regan and Noë (2001) claim that the feeling of 'seeing everything' is not so much an illusion, as it
     is a consequence of immediate availability: Everything is out there to see.

into the post-saccadic interval during which there is a gradual remapping from retinotopic to spatiotopic coordinates (Golomb et al., 2008; Mathôt & Theeuwes, 2010b).

Fourth, the visual system relies on the assumption of stability. That is, we perceive the world to be stable by default and substantial evidence to the contrary is required to break this assumption. This is related to the finding that a covert shift of attention precedes every eye movement (e.g., Kowler et al., 1995), allowing for an accurate preview of the saccade target. This preview is subsequently integrated with the post-saccadic percept of the saccade target, based on the assumption that the target has remained stable (Deubel & Schneider, 1994; McConkie & Currie, 1996). Not all items are equally important in the assumption of stability: Attention appears to determine which objects serve as a reference point. Like TSM, this theory does not make any claims about the underlying neurophysiology. However, one cannot help but wonder how this finding relates to remapping of RFs. It has been suggested that there is no direct relationship at all, but that both mechanisms are solutions to different problems: The assumption of stability underlies perceptual stability, whereas remapping is concerned with visually guided actions (Bays & Husain, 2007; also see Chapter v of this thesis). This is a plausible proposal, but an important avenue for future research will be to further investigate the relationship between remapping and the assumption of stability.

∞

III. REMAPPING AND VISUAL ATTENTION

Abstract — When attending an object in visual space, perception of the object remains stable despite frequent eye movements. It is assumed that visual stability is due to the process of remapping, in which retinotopically organised maps are updated to compensate for the retinal shifts caused by eye movements. Remapping is predictive when it starts before the actual eye movement. Until now, most evidence for predictive remapping has been obtained in single cell studies involving monkeys. Here we report that predictive remapping affects visual attention prior to an eye movement. Immediately following a saccade, we show that attention has partly shifted with the saccade (Experiment 1). Importantly, we show that remapping is predictive and affects the locus of attention prior to saccade execution (Experiments 2 and 3): before the saccade was executed there was attentional facilitation at the location which, after the saccade, would retinotopically match the attended location.

∞

Each time we make an eye movement the retinal projection of the world shifts dramatically. This stands in contrast with our conscious and stable perception of the world. It is believed that one way in which this perceptual stability is achieved is through a mechanism known as predictive remapping. Predictive remapping refers to the fact that in the interval in which an eye movement has been programmed but not yet executed, many visual neurons shift their receptive fields (RFs) from their current, pre-saccadic location to their post-saccadic location.

Predictive remapping was first described by Duhamel, Colby, and Goldberg (1992) in a classic study in which neural activity was recorded from the monkey lateral intraparietal area (LIP). They showed that just before the execution of a saccade some LIP-neurons become responsive to stimuli presented in their post-saccadic RF. They did this by presenting a stimulus outside of the neuron's RF and having the monkey make an eye movement to bring the stimulus into the neuron's RF. Even though no stimulus was present in their current (pre-saccadic) RF, just before the eye movement a subset of LIP-neurons

became active 'as if' the saccade had already been executed and had brought the stimulus into their (post-saccadic) RF. The effects of predictive remapping appear to be particularly strong in the PPC, but have been reported in a number of other oculomotor and visual areas as well, such as the extrastriate cortex, the frontal eye-fields and the superior colliculus (Nakamura & Colby, 2002; Umeno & Goldberg, 1997; Walker et al., 1995).

Recently, in a compelling psychophysical study, Melcher (2007) used the tilt-adaptation aftereffect (TAE) to demonstrate predictive remapping in human participants. TAE is a systematic bias in reporting the orientation of a tester stimulus, typically a slightly tilted grating, after being exposed for some time to a tilted adapter stimulus. After the presentation of the adapter stimulus, participants had to make an eye movement. The tester stimulus was presented at the location that after saccade execution would retinotopically match the adapter location. Crucially, TAE was found at this location even if the eye movement had not yet occurred, and therefore the tester and the adapter were not yet retinotopically matched. Predictive remapping in humans was also demonstrated in a recent ERP-study. Parks and Corballis (2008) showed that if a saccade carries a stimulus from one visual hemifield into the other, the corresponding interhemispheric shift of the stimulus' neurophysiological correlate occurs well before the saccade.
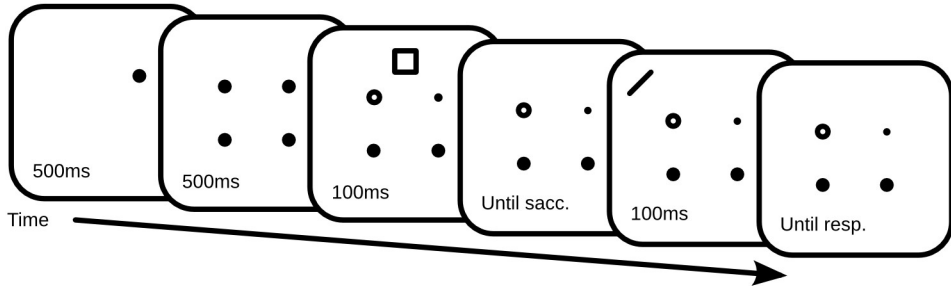
Remapping is considered to be a strong candidate mechanism for explaining why visual perception, or at least conscious experience thereof, is left largely undisturbed by eye movements. The intuitive notion that the brain contains a representation of the world in purely spatiotopic (world-centred) coordinates has received some support (e.g., Crespi et al., 2011; d' Avossa et al., 2007; Galletti et al., 1995), but is no longer favoured as the complete solution to the problem of visual stability (Colby & Goldberg, 1999; Mathôt & Theeuwes, 2011a; Wurtz, 2008). Instead, it is believed that visual information is represented largely in retinotopic (eye-centred) maps and is constantly remapped within those maps to compensate for saccade-induced retinal changes. Frequently, remapping occurs before the onset of a saccade in which case it is called predictive. One way of describing this process is that neurons receive a 'status report' on the location that will be brought into their RF by an impending saccade. This status report allows visual information to be preserved across saccades. The important role that remapping plays in visual stability is illustrated by a number of studies investigating remapping in human participants using functional magnetic resonance imaging (Medendorp, Goltz, Vilis, & Crawford, 2003; Merriam et al., 2003, 2007). In these studies, a visual stimulus was presented and subsequently extinguished, leaving a memory trace (residual neural activity). Next a saccade was executed such that the former stimulus location was carried

across the vertical meridian into the opposite visual field. The key finding in these studies was that, with the eye movement, the neural correlate of the memory trace shifts to the opposite hemisphere as well. Therefore, although the memory trace is represented in retinotopic coordinates, an eye movement does not cause the memory trace to be misaligned with the world. Rather, the memory trace is remapped onto a different set of neurons, such that the memory trace remains tied to the correct spatial location. Without the need for a spatiotopic representation of the world, this explains why we are able to integrate information presented at the same location before and after an eye movement (Ezzati et al., 2008; Melcher & Morrone, 2003; Wittenberg et al., 2008). Possibly, also the finding that in some cases people are very limited in their ability to integrate visual information across saccades (e.g., Irwin, 1991) can be explained in terms of remapping, or a lack thereof.

In a recent study, Golomb, Chun, and Mazer (2008) investigated how the locus of attention is affected by an eye movement. They aimed to answer the question what happens if attention is endogenously directed to a location, and subsequently the eyes move elsewhere while the attended location remains relevant to the task. While fixating, participants received a spatial cue and were instructed to hold the cued location in memory. It was assumed that keeping a location in memory requires the deployment of spatial attention to the memorised location (Awh & Jonides, 1998). After saccade execution Golomb and colleagues (2008) measured reaction times (RTs) to probes presented at the memorised (spatiotopic) location and the location that retinotopically matched the memorised location. RT facilitation was found at both locations. This suggests that the locus of attention is partly, but not entirely independent of eye position. More specifically, this suggests that the neural correlate of visual attention was partly remapped to compensate for the eye movement.

The present study investigated remapping of visual attention. We used the presentation of an irrelevant onset to manipulate attention exogenously (see e.g., Theeuwes, 1991; Yantis & Jonides, 1984). Experiment 1 was designed to determine whether exogenous attention is partly, but not completely remapped, as Golomb and colleagues (2008) reported in the case of endogenous attention. If so, we would expect attentional facilitation at the attended (spatiotopic) location as well as the location matching the attended location retinotopically. In Experiments 2 and 3 we tested whether remapping was predictive such that it would affect the allocation of attention prior to saccade execution. If so, we would expect attentional facilitation at the retinotopic location (now actually the 'future

a) Experiment 1: Actual retinotopic trial (schematic)
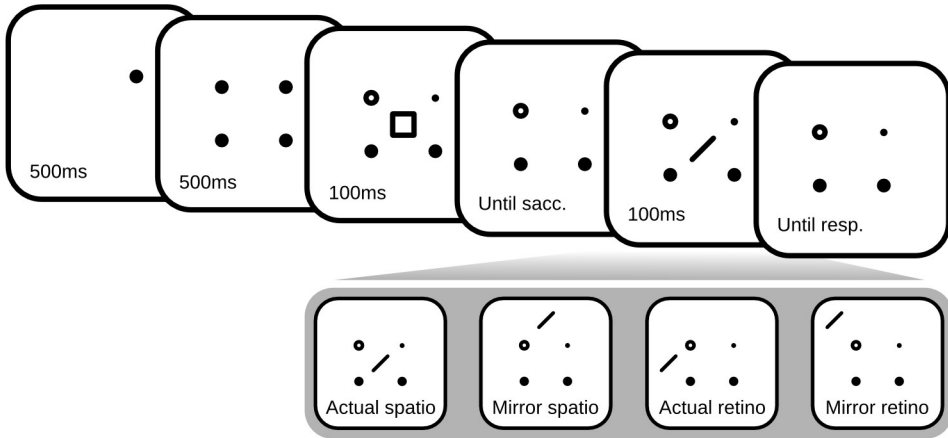

b) Experiment 1: Actual spatiotopic trial (schematic)

Figure iii.1. a and b) Two schematic example trials of Experiment 1. The saccade goal is denoted by the unfilled circle. a) An example of an actual retinotopic trial. b) An example of an actual spatiotopic trial. The grey box contains examples of probe positions in different conditions, in trials in which the onset was presented at the centre location.

retinotopic' location) after the eye movement has been cued, but before it has been executed.

Experiment 1

The paradigm was modelled after Golomb and colleagues (2008). Participants were instructed to execute a saccade. Simultaneously with the presentation of the saccade goal, an irrelevant abrupt onset was presented, which is known to capture attention. After the execution of the saccade, a probe stimulus (a tilted line-segment) appeared at one of four locations (the spatiotopic, retinotopic and two control locations). Participants made a speeded keypress response to indicate the orientation of the probe.

39

Method

Eighteen naive observers participated in the experiment. Eye movements were recorded using an Eyelink II (SR-research, Mississauga, ON, Canada). Each trial started with the presentation of a grey fixation dot on a black display at one of four possible locations (Figure iii.1). After 500ms, three additional and identical dots were presented, forming the corners of a 9.0° x 9.0° square. After another 500ms, the fixation dot reduced in size and one of the adjacent dots turned green, indicating that a saccade had to be made to that location. Participants did not know in advance to which location they had to execute a saccade. At the same time the dot turned green (the saccade cue) an onset stimulus (a 1.8° x 1.8° square) was presented for 100ms at one of two (given a certain fixation point and saccade cue) possible locations 6.4° from the initial fixation dot and the saccade cue. Participants were instructed to make a saccade to the green dot as quickly as possible. The saccade cue and the onset were presented simultaneously, because a delay between the onset and the saccade cue may lead to inhibition of the onset. Thirty milliseconds after the initiation of the saccade while the eyes were in motion a tilted grey line-segment (the probe) was presented for 100ms. We choose to present the probe in mid-flight (during saccadic suppression) rather than after the saccade to prevent the probe from capturing attention exogenously. The probe was presented sufficiently long for participants to observe it after they had re-fixated. The probe was presented with equal likelihood at one of four locations. The probe could be presented at the location that previously contained the onset (the Actual Spatiotopic location), at a location that retinotopically matched the onset location (the Actual Retinotopic location) or at one of two "Mirror" control locations. Participants made a speeded report of the probe orientation by pressing the "z"-key on a leftwards tilted line-segment (\) and the "/"-key on a rightwards tilted line-segment (/). The experiment consisted of 48 practice trials, followed by 256 experimental trials.

Results

Trials were discarded using the following criteria: gaze deviated more than 2° from the fixation point prior to the saccade cue (8.2%); the direction of the saccade deviated more than 22.5° from the straight line between the initial fixation point and the saccade goal (8.4%); saccade latency was either below 100ms or above 600ms (1.4%); RT was below 200ms or above 1000ms (2.3%). One participant was excluded due to loss of fixation (50% of the trials), one participant due to overly high saccade latencies (M = 449ms) and two participants due to a high proportion of misdirected saccades (25% and 34% respectively). Of the remaining participants, 79.7% of the trials were included in the analysis.
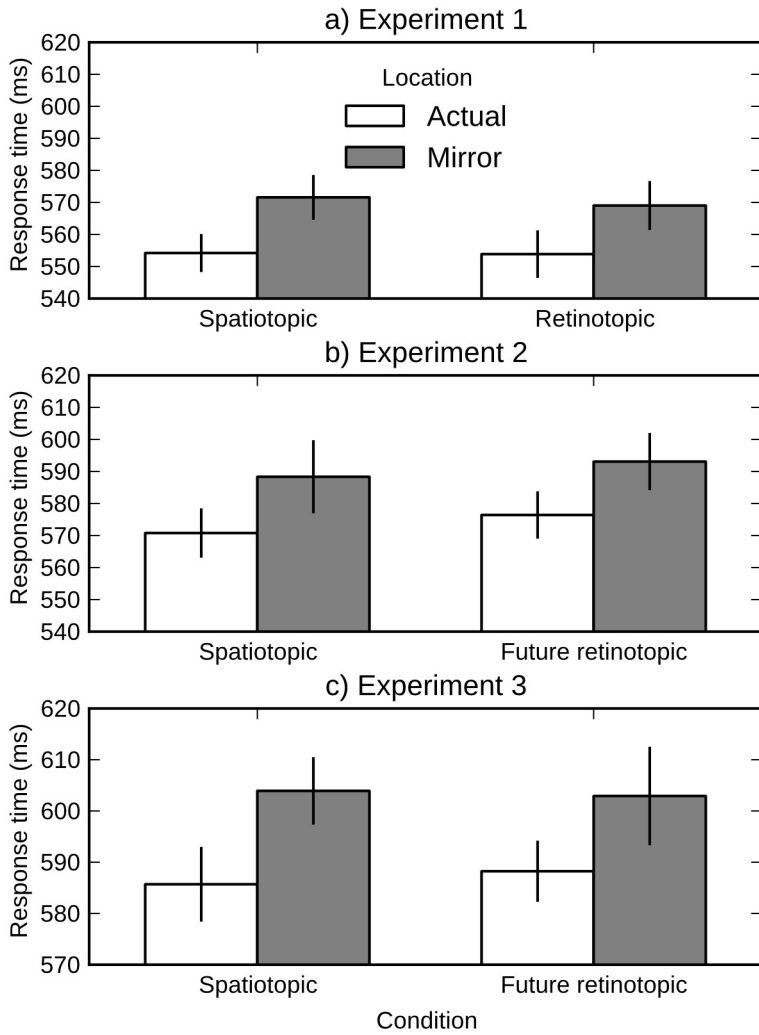
Figure iii.2. Results of Experiments 1 (a), 2 (b) and 3 (c). RTs are faster at the actual than at the mirror locations, indicating facilitation in both the (future) retinotopic and the spatiotopic condition. Error bars denote the 95% within-subject confidence interval (Cousineau, 2005).

A repeated measure analysis of variance (ANOVA), with Condition (Spatiotopic or Retinotopic) and Location (Actual or Mirror) as within-subject factors and mean correct RT as a dependent variable, revealed a main effect of Location, $F(1,13) = 11.5$, $p < .01$, (see Figure iii.2a). No other effects were found (all $F < 1$). Two-tailed paired samples t-tests revealed facilitation at the actual, $M = 554$ms, $SE = 20.9$, compared to the mirror spatiotopic location, $M = 572$ms, $SE = 23.2$, $t(13) = 3.1$, $p < .01$, and facilitation at the actual, $M = 554$ms, $SE = 21.1$, compared to the mirror retinotopic location, $M = 569$ms, $SE = 23.4$, $t(13) = 2.2$, $p < .05$). The same analysis with accuracy as dependent variable revealed a main effect of Location. $F(1,13) = 5.2$, $p < .05$ (Actual more accurate than Mirror), a marginally significant effect of Condition, $F(1,13) = 4.5$, $p < .1$ (Spatiotopic more accurate than Retinotopic) and no interaction between Location and Condition ($F < 1$). The average saccade latency was 270ms ($SE = 11.6$).

Discussion

The results indicate that, if attention is captured by an abrupt onset before a saccade is executed, immediately following the saccade attention resides at two locations: the original attended location and a second location that corresponds retinotopically to this location. These results resemble those of Golomb and colleagues (2008), showing partial remapping of visual attention, and extend these findings to exogenous attention.

Experiment 2

The aim of the second experiment was to investigate whether remapping of visual attention is predictive. In other words, we wanted to determine whether the focus of attention would shift slightly before the saccade was executed. The critical difference with Experiment 1 was that we presented the probe just before the eye movement.

Method

Twenty naive observers participated in the experiment. The method was the same as that of Experiment 1 except for the following differences. Eye movements were recorded using an Eyelink 1000 (SR Research, Mississauga, ON, Canada). The line-segment was presented at a fixed interval of 125ms after the presentation of the saccade cue. The onset was presented for 50ms and the probe was presented for 75ms, to assure that on most trials all stimuli were presented before the saccade was initiated. The retinotopic location is now referred to as the 'future-retinotopic' location.

Results

Trials were filtered using the same criteria as in Experiment 1: gaze deviation (11.0%), saccade direction (14.0%), saccade latency (1.3%) and RT (2.0%). In addition, trials in which the eyes arrived at the saccade target before the offset of the line-segment were discarded (7.9%). One participant was excluded due to anticipatory eye movements on 28% of the trials. Of the remaining participants, 63.9% of the trials were included in the analysis.

A repeated measures ANOVA, using Condition (Spatiotopic or Future Retinotopic) and Location (Actual or Mirror) as within-subject factors and mean correct RT as a dependent variable, revealed a main effect of Location, $F(1,18) = 8.0$, $p < .05$ (see Figure iii.2b). No other effects were found (all $F < 1$). Paired samples t-tests revealed facilitation at the actual, $M = 571$ms, $SE = 18.3$, compared to the mirror spatiotopic location, $M = 588$ms, $SE = 19.7$, $t(18) = 2.2$, $p < .05$, and facilitation at the actual, $M = 576$ms, $SE = 19.4$, compared to the mirror future retinotopic location, $M = 593$, $SE = 20.5$, $t(18) = 2.5$, $p < .05$. The same analysis with accuracy as dependent variable revealed a main effect of Condition, $F(1,18) = 10.5$, $p < .01$ (Spatiotopic more accurate than Future Retinotopic). No other main effects were found. The average saccade latency was 269ms ($SE = 13.7$).

Discussion

The results indicate that predictive remapping affects the locus of attention in the interval preceding a saccade. Before the saccade was executed, attentional facilitation was observed at the 'future retinotopic' location: the location that retinotopically matched the attended location after the eye movement.

Experiment 3

Even though Experiment 2 provides clear evidence for predictive remapping, there is one caveat. In Experiment 2 the Actual and Mirror locations were always presented in opposite visual fields, separated by the horizontal or the vertical meridian. Therefore, the future-retinotopic facilitation could have been due to a spreading of attention from the onset location to other areas within the same visual field quadrant (e.g., Rizzolatti et al., 1987). Since spreading of attention across the horizontal or vertical meridian is less pronounced, one may obtain RT differences that have nothing to do with predictive remapping. To address this issue, we adapted the paradigm to allow saccades in every direction. Therefore, the future-retinotopic location did not always fall in the same visual quadrant as the onset (see Figure iii.3).
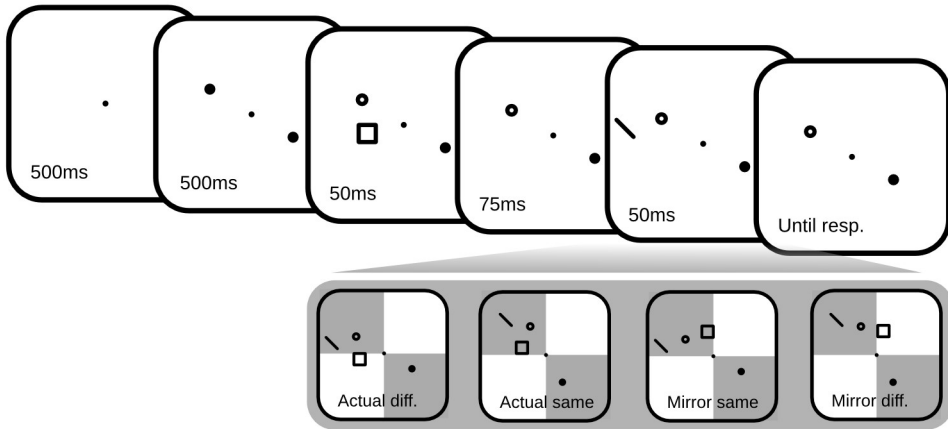
Figure iii.3. A schematic example trial of Experiment 3 in the actual future-retinopic condition. In this example, the probe and the onset are presented in different visual quadrants, in this case on opposite sides of the horizontal meridian, but it could be on opposite sides of the vertical meridian as well. The grey box contains example stimulus configurations for actual and mirror future-retinotopic trials. The probe and the onset could be presented in the same or in different visual quadrants; analysis revealed that there was no effect of visual quadrant (see the Results section of Experiment 3). Visual quadrants are marked by shades of grey for convenience. Actual and mirror spatiotopic trials were included in Experiment 3 as well, but they are not depicted here.

Method

   9 naive observers participated in the experiment. The method was similar to that of Experiment 2 except for the following differences. The initial fixation dot was always presented at the centre of the display. There were two potential saccade targets, 7.2° from and on opposite sides of the fixation point (Figure iii.3). Given these constraints, the position of the potential saccade targets was random. All stimuli were scaled to 80% of the size used in Experiments 1 and 2. The presentation duration of the probe was reduced to 50ms, to reduce the number of trials on which the probe was still visible upon arrival at the saccade target. There were 384 experimental trials.

Results

   Using the same criteria as in Experiment 2, trials were filtered on gaze deviation (9.7%), saccade direction (13.3%), saccade latency (1.9%), RT (3.2%) and premature arrival at the saccade target (0%). In total, 71.8% of the trials were included in the analysis.

   We performed the same analysis as in Experiment 2 and obtained qualitatively identical results: RT data revealed a main effect of location, $F(1,8) = 13.3$, $p < .01$ (see Figure iii.2c) and facilitation at both the spatiotopic (actual: $M = 586$ms, $SE = 22.2$; mirror: $M = 604$ms,

SE = 23.3), *t*(8) = 4.0, *p* < .01, and the future-retinotopic location (actual: M = 588ms, SE = 22.2; mirror: *M* = 603ms, *SE* = 22.5), *t*(8) = 2.5, *p* < .05. Accuracy data revealed a marginally significant effect of Condition, *F*(1,8) = 4.8, *p* < .1. The average saccade latency was 250ms (*SE* = 17.1).

We performed an additional analysis on trials in the Future-retinotopic condition and included as a factor whether the onset and the probe were presented in the same or in different visual quadrants (see Figure iii.3). A repeated measures ANOVA using Location and Quadrant as within-subject factors and mean correct RT as dependent variable revealed a main effect of Location, *F*(1,8) = 5.8, *p* < .05, but no other effects (all *F* < 1), indicating that there was no meridian effect (same quadrant actual, *M* = 588ms; mirror, *M* = 609ms; different quadrant actual, *M* = 591ms; mirror *M* = 603ms).

Discussion

In Experiment 3 we confirmed that predictive remapping affects the locus of attention in the interval preceding a saccade. In addition, we ruled out a meridian effect as an alternative explanation.

General discussion

The present study clearly shows that predictive remapping affects the allocation of attention prior to an eye movement The presentation of a brief onset attracted attention, resulting in a temporary increase in baseline activity of neurons whose RFs overlap the attended location (for a review, see Reynolds & Chelazzi, 2004). When during this short interval a probe is presented at the attended location, this allows for a faster and more accurate report of the probe identity. Most of the visual system is retinotopically organised. In Experiment 1, this resulted in attentional facilitation at the location that retinotopically matched the onset. However, neural responses are often remapped to compensate for eye movements. Similar to the remapping of a memory trace (e.g., Merriam et al., 2003), the activation elicited by the onset was transferred to a different population of neurons, which after the saccade had RFs encompassing the original onset location. This resulted in attentional facilitation at the original onset location. Similar findings showing both retinotopic and spatiotopic integration across saccades have been reported for a variety of phenomena (Ezzati et al., 2008; Golomb et al., 2008; Melcher, 2005; Wittenberg et al., 2008). Crucially, in Experiments 2 and 3 we showed that the locus of attention partly shifts in the direction of the saccade, prior to the eye movement. This can be explained by assuming that the neurons which have been activated by the presentation of the onset remap predictively: they exhibit an anticipatory RF shift in the direction of the saccade.

Therefore, they will respond to the presentation of the probe at the future-retinotopic location, allowing the probe to 'ride the wave' of the onset.

Remapping is believed to be crucial in maintaining visual stability (Wurtz, 2008). Most neurons have RFs that are anchored to the retina. Therefore, after an eye movement they are exposed to a different part of the visual scene. However, rather than perception starting anew after every eye movement, it seems that neurons receive a 'status report' on the location that will be brought into their RF by the eye movement. Frequently, this process starts in the interval preceding saccade execution, in which case it is referred to as predictive (Duhamel et al., 1992; Melcher, 2007). Here we report that exogenous visual attention is remapped, but only partly. Therefore, after an eye movement attention is allocated at two locations: the original locus of attention and the location that retinotopically matches the original locus of attention. Importantly, we also show that predictive remapping causes the locus of attention to partly shift in the direction of an eye movement prior to saccade execution.

<div align="center">∞</div>

IV. Remapping and Inhibition of Return

Abstract — Here we report that immediately following the execution of an eye movement oculomotor inhibition of return resides in retinotopic (eye-centred) coordinates. At longer post-saccadic intervals inhibition resides in spatiotopic (world-centred) coordinates. These results are explained in terms of peri-saccadic remapping. In the interval surrounding an eye movement, information is remapped within retinotopic maps to compensate for the retinal displacement. Because remapping is not an instantaneous process, a fast, but gradual transfer of inhibition of return from retinotopic to spatiotopic coordinates can be observed in the post-saccadic interval. The observation that visual stability is preserved in inhibition of return is consistent with its function as a 'foraging facilitator', which requires locations to be inhibited across multiple eye movements. The current results support the notion that the visual system is retinotopically organised and that the appearance of a spatiotopic organisation is due to remapping of visual information to compensate for eye movements.

∞

We perceive the world as stable and are capable of making accurate, visually guided movements. This is remarkable considering that visual perception relies on input from the retina which, because of eye movements, is unstable. The apparent conflict between the lack of stability in visual input and the stable nature of conscious visual perception is often referred to as the problem of visual stability.

An intuitively appealing solution to this problem is to assume that only early visual areas are affected by eye movements. It is an old idea, frequently attributed to Helmholtz or even pre-Socratic philosophers, that information about eye position is crucial in maintaining visual stability (Grüsser, 1986). The spatiotopic hypothesis posits that action and conscious perception are based on a world-centred (spatiotopic) representation that is constructed by combining gaze-centred (retinotopic) maps with eye position information. However, there is little empirical support for this hypothesis. The most convincing

evidence so far has been provided by an fMRI study showing spatiotopic responses in the human middle temporal cortex (d' Avossa et al., 2007), but this finding has recently been contested (Gardner et al., 2008; but see Crespi et al., 2011). The situation is somewhat different in multimodal parietal areas, where gaze modulated visual receptive fields (RFs) are frequently observed (Duhamel et al., 1997; Galletti, Battaglini, & Fattori, 1993; Galletti et al., 1995). However, there is ongoing debate about how these RFs are best characterised (Mullette-Gillman, Cohen, & Groh, 2005).

The fact that the visual system is largely retinotopically organised has caused many researchers to look beyond spatiotopy for mechanisms underlying visual stability (for recent reviews, see Mathôt & Theeuwes, 2011a; Wurtz, 2008). A mechanism which has received considerable interest is remapping of RFs (Duhamel et al., 1992). According to the retinotopic (or remapping) hypothesis information is transferred within retinotopic maps to compensate for eye movements. Consider, for example, a stimulus that is presented briefly and then extinguished, leaving some residual neural activity. If an eye movement is executed such that the former stimulus location is carried across the vertical meridian, there is a corresponding shift of neural activity from one hemisphere to the other, which can be measured using fMRI (Merriam et al., 2003). The effect of this inter-hemispheric remapping is that the memory trace does not become misaligned with the world after the eye movement. Therefore, in many ways remapping gives the appearance of a spatiotopically organised visual system, while relying on retinotopic maps.

A prime example of a phenomenon in which maintaining visual stability is crucial is Inhibition of Return (IOR). IOR refers to the fact that people tend not to revisit recently attended locations, either with their eyes or by a covert shift of attention (Posner & Cohen, 1984). IOR is often described as a 'foraging facilitator', because it facilitates visual search by preventing us from visiting the same locations over and over again, while neglecting most of the visual field (Klein, 1988). Because IOR is a sustained effect, spanning multiple eye movements, it would be maladaptive if it were tied to retinotopic coordinates. Indeed, a number of studies have shown that IOR resides in spatiotopic coordinates (Maylor & Hockey, 1985; Posner & Cohen, 1984; but see Abrams & Pratt, 2000) and can be maintained across multiple eye movements (Klein & MacInnes, 1999; but see Ludwig et al., 2009). More recently, studies have shown that IOR has both a spatiotopic and retinotopic component, but that the spatiotopic component can be disrupted by parietal lesions (Sapir, Hayes, Henik, Danziger, & Rafal, 2004) or the application of TMS to the right parietal cortex (van Koningsbruggen et al., 2010). These findings clearly show that

stability is somehow preserved in IOR, presumably by a process that involves the parietal cortex. However, the underlying mechanisms are still unclear.

In the present study we investigated whether the spatiotemporal properties of IOR are consistent with the hypothesis that IOR relies on retinotopic maps and is remapped gradually in the peri-saccadic interval. The retinotopic hypothesis predicts that a particular visual phenomenon may appear to be spatiotopic or retinotopic depending on the time of measurement, because remapping is a process that takes time (Kusunoki & Goldberg, 2003): Remapping starts well before the onset of a saccade (Duhamel et al., 1992), but visual stability is not fully restored until sometime after a saccade (e.g., Hamker et al., 2008; Morris et al., 2012). A recent study by Golomb and colleagues (2008) is of particular relevance here. Golomb and colleagues investigated how the reference frame of attentional facilitation, which typically precedes the inhibitory phase investigated here, changes in the post-saccadic interval. In one experiment they showed retinotopic facilitation directly after a saccade and spatiotopic facilitation at longer intervals. However, in a second experiment, in which participants were asked to attend to a location 'relative to their eyes', no spatiotopic facilitation was found. This led the authors to conclude that the reference frame of visual attention is retinotopic, unless remapping is required to perform the task at hand. While we agree that visual attention relies on a retinotopically organised visual system, we have previously shown that the locus of attention is also remapped when attention is manipulated using a task-irrelevant onset stimulus (Mathôt & Theeuwes, 2010a).

The aim of the present study was to investigate whether the reference frame of IOR changes in the post-saccadic interval, as has been shown for attentional facilitation (Golomb et al., 2008). We predicted that IOR is predominantly retinotopic when measured just after an eye movement. If the delay between the eye movement and the time of measurement is relatively long, we expected IOR to be predominantly spatiotopic. To test this hypothesis we briefly presented an onset stimulus in order to exogenously attract attention, eliciting IOR at longer intervals (Posner & Cohen, 1984). Subsequently, participants made two eye movements. The first eye movement allowed us to dissociate spatiotopic and retinotopic coordinates. The latency of the second eye movement allowed us to determine the strength of IOR at the retinotopic and spatiotopic locations.

Method

Thirteen naive observers and one of the authors (SM), all between the ages of 18 and 27, participated in the experiment. All participants reported normal or corrected visual acuity. Eye movements were recorded using an Eyelink 1000 (SR-Research, Mississauga, ON, Canada), a video-based eye-tracker sampling at 1000Hz.
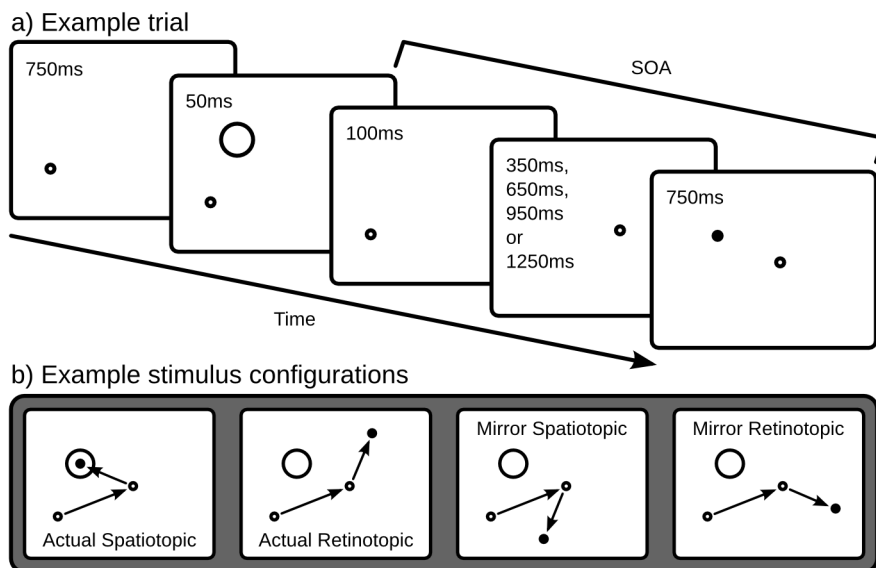
Figure iv.1. a) A schematic example trial in the 'actual spatiotopic' condition. Participants made two eye movements. The first eye movement matched the displacement of the fixation dot (small empty circle) and allowed us to dissociate retinotopic and spatiotopic coordinates. The second eye movement was to the green saccade target (small filled circle). We manipulated attention by briefly presenting an onset stimulus (large empty circle). b) The four possible stimulus configurations and corresponding eye movements (arrows). The saccade target could be presented at the onset location (actual spatiotopic), the location that retinotopically matched the onset location (actual retinotopic) or at one of two control locations (mirror spatiotopic and mirror retinotopic). Stimulus arrangements were randomly rotated and positioned on the display.

Before the start of each trial a grey cross was presented against a dark background. Drift-correction was executed automatically when participants fixated this cross (except for in the first trial of each block in which a keypress was required), after which the trial was initiated. Each trial started with the presentation of a grey fixation dot (see Figure iv.1a). After 900ms, the fixation dot was displaced 6° to a random location. Participants followed the displacement with their eyes. After a variable interval (350ms, 650ms, 950ms or 1250ms) a saccade target (a green dot) was presented 4.2° from the fixation dot for 750ms. Participants made a saccade to the target as quickly as possible. The location of the saccade target was used as the starting location for the next trial.

Attention was manipulated by the presentation of an onset stimulus (an empty circle, *d* = 0.5°) for 50ms, 150ms prior to the displacement of the fixation dot. The onset stimulus and the green saccade target were presented at the same spatial location (actual spatiotopic; see Figure iv.1b), the same location relative to the fixation dot (actual retinotopic), or in

one of two comparable control arrangements (mirror spatiotopic and mirror retinotopic). Stimulus arrangements were rotated randomly so that there was a mixture of trials in which the onset location needed to be remapped across the vertical vertical meridian and trials in which this was not the case.

If gaze deviated more than 2° from the expected position during the experiment, the participant received auditory feedback and the trial was continued, but discarded. The experiment consisted of 30 practice trials, followed by 640 experimental trials.

Results

Trials in which gaze deviated more than 1° from the initial fixation dot, more than 2° from the displaced fixation dot, more than 2° from the saccade target, or in which the required eye movements were not executed, were excluded (22.2%). The primary reason for gaze error was oculomotor capture by the onset, which accounted for about half of these errors, consistent with previous studies (Theeuwes, Kramer, Hahn, & Irwin, 1998). Trials in which the latency of the second saccade (to the saccade target) was less than 50ms or more than 500ms were excluded (4.7%). The target of the second saccade was never presented prior to the initiation of the first saccade. In total, 74.2% of the trials were included in the analysis.

A repeated measures analysis of variance (ANOVA) with condition (spatiotopic or retinotopic), location (actual or mirror) and stimulus onset asynchrony (SOA; the interval between the presentation of the onset stimulus and the saccade target; 500ms, 800ms, 1100ms or 1400ms) as independent variables and the latency of the second saccade (to the saccade target) as dependent variable revealed the following effects (see Figure iv.2 and Table iv.1): a main effect of location ($F(1, 13) = 22.3$, $p < .001$, $\eta_p^2 = .63$), reflecting an overall IOR effect; a main effect of condition ($F(1, 13) = 8.0$, $p < .05$, $\eta_p^2 = .38$), reflecting lower latencies in the retinotopic, relative to the spatiotopic condition; a main effect of SOA ($F(3, 13) = 10.0$, $p < .001$, $\eta_p^2 = .44$), reflecting lower latencies at higher SOAs; an interaction between condition and SOA ($F(3, 39) = 4.15$, $p < .05$, $\eta_p^2 = .24$), reflecting a more pronounced latency decrease with increasing SOA for the spatiotopic condition. Crucially, we also found a three-way interaction between condition, location and SOA ($F(3, 39) = 3.4$, $p < .05$, $\eta_p^2 = .21$), reflecting the different time course for retinotopic and spatiotopic IOR.

Two tailed paired samples t-tests revealed IOR (as measured by increased saccade latencies to the actual, relative to the mirror locations) for the retinotopic condition at 500ms SOA (actual: $M = 260$ms, $SE = 12.7$; mirror: $M = 236$ms, $SE = 15.5$; $t(13) = 3.3$, $p < .01$), and for the spatiotopic condition at the 1100ms (actual: $M = 230$ms, $SE = 10.8$;

Figure iv.2. The results of the experiment. 500ms after the presentation of the onset (SOA) and on average 73ms after the end of the first saccade (FOA) there was retinotopic IOR. At 1100ms SOA (634ms FOA) and 1400ms SOA (922ms FOA) there was spatiotopic IOR. The FOA standard error is provided within parentheses. IOR was measured as an increased saccade latency of the second saccade to the actual, relative to the mirror locations (see Figure iv.1b). Asterisks denote significant IOR (†p < .1, *p < .05, **p < .01, ***p < .005). Error bars represent a 95% within-subject confidence interval (Cousineau, 2005). Overlapping error bars have been removed for clarity.

mirror: M = 215ms, SE = 10.6; $t(13)$ = 3.6, p < .005) and 1400ms SOAs (actual: M = 226ms, SE = 9.4; mirror: M = 217ms, SE = 8.1; $t(13)$ = 2.8, p < .05). There was a marginally significant spatiotopic IOR-effect at the 800ms SOA (actual: M = 253ms, SE = 10.7; mirror: M = 240ms, SE = 12.7; $t(13)$ = 2.2, p = .05).

Because our main goal was to investigate how the reference frame of IOR changes in the post-saccadic interval, we determined the time between the end of the first saccade and the presentation of the target for the second saccade. The average fixation onset asynchrony (FOA) in the 500ms SOA was 73ms (*SE* = 9.3), 346ms (*SE* = 11.8) in the 800ms SOA, 634ms (*SE* = 12.1) in the 1100ms SOA and 922 (*SE* = 14.6) in the 1400ms SOA.

Table iv.1. An Overview of Reaction Times (ms) per Condition

SOA (ms)	Spatiotopic			Retinotopic		
	Actual	Mirror	IOR	Actual	Mirror	IOR
500	272	265	7	260	236	**24
800	253	240	†13	247	238	9
1100	230	215	***15	225	219	6
1400	226	217	*9	217	214	3

Note. †*p* < .1, **p* < .05, ***p* < .01, ****p* < .005

Discussion

The present study shows that the spatiotemporal properties of IOR are consistent with the idea that IOR relies on retinotopic maps and is remapped gradually in the peri-saccadic interval. Only at the shortest interval, on average 73ms after the eye movement (500ms after the presentation of the onset), there was IOR at the location which retinotopically matched the onset location. IOR at the original onset (spatiotopic) location was found at longer post-saccadic intervals. This finding illustrates that remapping of IOR is a gradual process that persists into the post-saccadic interval.

The present study complements existing literature in a number of ways. Most support for remapping has been obtained in neurophysiological (e.g., Duhamel et al., 1992; Kusunoki & Goldberg, 2003) and neuroimaging studies (e.g., Merriam et al., 2003; Parks & Corballis, 2008). Studies which have employed behavioural methods have generally focused on the pre-saccadic interval (Hunt & Cavanagh, 2009; Mathôt & Theeuwes, 2010a; Melcher, 2007) or have addressed the reference frame of visual phenomena without directly investigating their dynamic nature (e.g., Ezzati et al., 2008; Knapen et al., 2009). Here we show that the reference frame of IOR is highly dynamic, as has been shown previously for attentional facilitation (Golomb et al., 2008).

Recently, Pertzov and colleagues (Pertzov et al., 2010) reported strong spatiotopic IOR immediately after a saccade. This finding is surprising given the study by Golomb et al. (2008) and the present study, which suggest that remapping of attentional effects

(facilitation as well as IOR) is not yet complete at the time a saccade is finished. This apparent discrepancy warrants further scrutiny.

Given the assumption that remapping is predictive—that is, occurring before, instead of after a saccade—the current findings may be surprising. However, we believe that the present study complements rather than challenges the notion of predictive remapping. Although we did not find significant spatiotopic IOR at the early post-saccadic intervals, we do not contest that early spatiotopic effects can be observed in different experimental paradigms (e.g., Mathôt & Theeuwes, 2010a). The crucial finding here is that even after an eye movement has ended it takes some time for spatiotopic IOR to be fully restored (see also Golomb et al., 2008). Therefore, we prefer to use the term remapping not only for pre-saccadic processes, but for all peri-saccadic processes that allow for a translation from retinotopic to spatiotopic coordinates.

Given the presumed role of IOR as a 'foraging facilitator' and the fact that IOR typically spans multiple eye movements, one would not expect IOR to be purely retinotopic: It would be maladaptive to inhibit a novel location which happens to be in the same retinal position as a previously visited location. Consistent with this functional view, previous studies have shown that IOR resides in spatiotopic coordinates (Klein & MacInnes, 1999; Maylor & Hockey, 1985; Posner & Cohen, 1984). However, the present study shows that spatiotopy is not a fundamental property of IOR, but emerges through the active process of remapping. An alternative explanation would be that the current findings do not reflect remapping of IOR per se, but rather remapping of the locus of attention, with the transition from facilitation to inhibition occurring after the first saccade. However, this is highly unlikely given the short interval (on average 73ms) between the end of the first saccade and the earliest time of measurement, since the transition from facilitation to inhibition takes at least 200ms (e.g., Posner & Cohen, 1984).

In summary, the present study shows that the locus of IOR is remapped in the post-saccadic interval. After an eye movement there is a narrow temporal window during which visual stability has not yet been fully restored. During this period, the retinotopic nature of the visual system can be exposed. At longer post-saccadic intervals, remapping has largely compensated for the eye movement, thus giving, at a behavioural level, the appearance of spatiotopy, while relying on a retinotopically organised visual system.

Appendix: Comment on Hilchey et al. (in press)[9]

When attention is directed to a location and subsequently withdrawn, people are slower to respond manually to stimuli presented at that location (Posner & Cohen, 1984), and slower to make an eye movement to that location (Abrams & Dobkin, 1994; Klein & MacInnes, 1999; for a review, see Klein, 2000). This phenomenon is called inhibition of return (IOR), and is often interpreted as a 'been there, done that' mechanism, or foraging facilitator, that prevents us from revisiting the same locations over and over again. IOR thus facilitates visual search (Klein, 1988) and, presumably, our daily life interactions with our visual environment. Assuming that IOR serves this purpose, one would expect the locus of IOR to be tied to spatiotopic (world-centred) coordinates: Regardless of the current position of gaze, the same location 'out there' should be inhibited. It would make little sense for IOR to be tied to retinotopic (gaze-centred) coordinates, thus arbitrarily shifting around with each eye movement. This line of reasoning is supported by studies showing that IOR spans multiple eye movements (Klein & MacInnes, 1999; but see Ludwig et al., 2009) and is indeed, at least in part, tied to a spatiotopic frame of reference (Hilchey, Klein, Satel, & Wang, in press; Mathôt & Theeuwes, 2010b; Maylor & Hockey, 1985; Pertzov et al., 2010; Posner & Cohen, 1984; Sapir et al., 2004; van Koningsbruggen et al., 2010). In other words, IOR is largely robust to eye movements—as it should be.

We recently investigated the reference frame of IOR in more detail (Mathôt & Theeuwes, 2010b). Consistent with previous reports, we found that IOR has both a retinotopic and a spatiotopic component (Hilchey et al., in press; Pertzov et al., 2010; Sapir et al., 2004). However, crucially, we also found a difference in the time-course of retinotopic and spatiotopic IOR. Retinotopic IOR was strongest immediately after a saccadic eye movement (at short post-saccadic delays), but dissipated quickly. In contrast, spatiotopic IOR was more sustained and even appeared to increase at longer post-saccadic delays (this latter effect was not as pronounced, but see e.g. Golomb et al., 2011). We interpreted this as reflecting a 'remapping' mechanism that restores spatiotopic IOR after an eye movement, but leaves a transient retinotopic trace (cf. Golomb et al., 2008; Mathôt & Theeuwes, 2010a). This result is important because it helps to understand how spatiotopic IOR, and presumably other spatiotopic visual phenomena as well, emerge from a visual system that is by-and-large retinotopically organised (Golomb & Kanwisher, in press; Talbot & Marshall, 1941; for recent reviews, see Mathôt & Theeuwes, 2011a; Wurtz, 2008).

---

9   This appendix contains a commentary on a recent paper by Hilchey and colleagues (in press), and was not part of the original publication as it appeared in *Psychological Science*.

However, in a recent paper, Hilchey, Klein, Satel, and Wang (in press) re-analysed our dataset (from Mathôt & Theeuwes, 2010b available from the first author's website) and came to an altogether different conclusion. Specifically, they excluded all saccadic responses that were faster than 100ms (1.4 standard deviations below the mean), whereas we had used a more conservative lower bound of 50ms (2.0 standard deviations below the mean). According to Hilchey and colleagues (in press), a lower bound of 100ms was more appropriate, because they considered faster saccades to be of little theoretical significance. Furthermore, visual inspection of the saccade latency distribution appeared to indicate that saccades faster than 100ms were outliers (see Fig A1 from Hilchey et al., in press). After filtering the data based on this new exclusion criterion, Hilchey and colleagues (in press) found that the retinotopic component of IOR was substantially attenuated, particularly for the shortest SOA at which this component was strongest. Consequently, the difference in time-course, which was central to our story, was no longer obvious. Based on this, Hilchey and colleagues (in press) concluded that our claim of early retinotopic and late spatiotopic IOR was based on an artefactual result, due to the inclusion of theoretically uninteresting "impossibly fast" saccades. They were strengthened in this conclusion by the fact that their re-analysis brought our results more in line with their own results (Hilchey et al., in press) and those of Pertzov and colleagues (2010).

The aim of the present paper is two-fold. First, we want to share an elaborate re-analysis of that same dataset (from Mathôt & Theeuwes, 2010b). We have characterised a previously unrecognised, but crucial, effect of saccade latency (saccadic response time) on the reference frame of IOR. Specifically, retinotopic IOR is most pronounced for fast saccadic responses, whereas spatiotopic IOR is most pronounced for slow responses. We will show that by taking saccade latency into account, the dissociation between retinotopic and spatiotopic IOR becomes even more conspicuous—Unlike suggested by Hilchey and colleagues (in press), this dissociation is a robust phenomenon. However, our re-analysis does call for a modification, or rather extension, of our original interpretation (cf. Mathôt & Theeuwes, 2010b): IOR is predominantly retinotopic immediately after an eye movement, *particularly for fast saccadic responses*, and predominantly spatiotopic at longer post-saccadic delays, *particularly for slow saccadic responses*. Thus, the dominant reference frame of IOR is determined by both the speed of the saccadic response (saccade latency) and the time at which the saccade target is presented relative to a preceding eye movement (post-saccadic delay).

Second, we aim to refocus the debate on the central issue, which is one of timing. Hilchey and colleagues (in press) did not include a timing manipulation in their own

experiment, in which the saccade target was always presented immediately following an eye movement. However, the central issue is emphatically not the absolute strength of retinotopic and spatiotopic IOR at any single point in time. Although we did not ourselves find significant spatiotopic IOR immediately after a saccade, we explicitly avoided the implication that early spatiotopic effects do not exist (Mathôt & Theeuwes, 2010b, p. 1797). Instead, the central issue is the differential time-course of retinotopic and spatiotopic IOR: The finding that (according to our results) retinotopic IOR decays rapidly after a saccadic eye movement, whereas spatiotopic IOR is sustained and appears to increase somewhat with longer post-saccadic delays. In our view, this differential time-course is evidence for what we have referred to as 'gradual remapping', or the notion that visual stability is not fully restored until sometime (±150ms) after an eye movement. Here we re-assert this claim, and extend it by describing the (complementary) effect of saccade latency on the reference frame of IOR.

Methods and analysis

We present a novel analysis of a previously reported dataset. The full experimental design and data pre-processing procedure are described in Mathôt and Theeuwes (2010b). Trials were divided into bins based on saccade latency, separately for each combination of subject ($N = 14$), coordinate system (spatiotopic, retinotopic), and saccade target location (actual, mirror). Separate analyses were conducted across all SOAs (using 25 bins; Figure iv.3a), and for each SOA separately (using 10 bins; Figure iv.3b)[10]. For each saccade latency bin, we calculated Bayes factors (Bfs), based on the *retinotopic IOR – spatiotopic IOR* contrast, and the *spatiotopic IOR – retinotopic IOR* contrast, to investigate whether there was a difference between retinotopic and spatiotopic IOR and, if so, in which direction. We assumed a uniform distribution, a lower bound of 0ms and an upper bound of 50ms. These bounds were chosen because the effect of IOR tends to be in the 0 to 50ms range (e.g., Hilchey et al., in press; Klein, 2000; Mathôt & Theeuwes, 2010b; Pertzov et al., 2010; Posner & Cohen, 1984). (The effects reported here are not crucially dependent on the exact bounds and number of bins.) We considered only series of at least two consecutive bins that showed "substantial evidence" for a difference in one contrast (Bf > 3; cf. Jeffreys, 1961; reproduced in Wetzels et al., 2011).

---

10   Because of the large variation of saccade latencies between subjects and, to a lesser extent, between conditions, the distributions of saccade latencies in adjacent bins overlap (Figure iv.3a,b). Therefore, it is difficult to pinpoint precisely at which saccade latencies retinotopic and spatiotopic IOR become dominant. Looking at data from individual subjects (not shown), it is clear that there is substantial variation in this respect, although the general effect of saccade latency is evident in almost every subject.

Figure iv.3. a) The effect of saccade latency on the strength of retinotopic and spatiotopic IOR across all SOAs. For fast saccades, retinotopic IOR is stronger than spatiotopic IOR. For slow saccades, spatiotopic IOR is stronger than retinotopic IOR. b) The effect of saccade latency and SOA/ post-saccadic delay on retinotopic and spatiotopic IOR. The most prominent result is that retinotopic IOR is largest for fast saccades in the shortest SOA (73ms post-saccadic delay), whereas spatiotopic IOR is largest for slow saccades in the 1100ms SOA (634ms post-saccadic delay). Error bars indicate within-subject standard errors (Cousineau, 2005). Shaded areas indicate evidence for a difference between retinotopic and spatiotopic IOR (see text).

The analysis employed here is unconventional, but appropriate for two reasons. First, when conducting a large number of post-hoc comparisons, a Bayesian approach is preferable to null hypothesis testing (Dienes, 2011). Second, our criteria were conservative, because "substantial evidence" following Jeffrey's (1961) criterion is more stringent than the commonly used alpha level of .05 (Wetzels et al., 2011), and we furthermore considered only series of at least two bins that showed substantial evidence for a difference in the same direction.

Discussion and results

It is clear that there is a large effect of saccade latency on the reference frame of IOR when considered across all SOAs (Figure iv.3a). For fast saccadic responses (approx. <150ms) retinotopic IOR is larger than spatiotopic IOR, whereas this pattern reverses for slow saccades (approx. >250ms)[10]. For saccades of intermediate latency there is no clear difference between spatiotopic and retinotopic IOR. The effect of saccade latency may explain why recent investigations of the reference frame of IOR have led to seemingly conflicting results (Hilchey et al., in press; Mathôt & Theeuwes, 2010b; Pertzov et al., 2010). More specifically, this explains why early retinotopic IOR may be attenuated (relative to our findings) when strict trial exclusion criteria are applied (<100ms, cf. Hilchey et al., in press; <150ms, cf. Pertzov et al., 2010) or when there is a low incidence of very fast saccades (as reported by Hilchey et al., in press).

An obvious question is whether the effects of saccade latency and post-saccadic delay (manipulated through SOA in our experiment) have been confounded in the past. Could the effect of post-saccadic delay that we previously reported (Mathôt & Theeuwes, 2010b) have been an indirect effect of saccade latency? However, looking at Figure iv.3b, it appears that both effects are distinct (see also Table 1 from Mathôt & Theeuwes, 2010b), because it is not the case that the strongest retinotopic IOR is observed in the condition that elicited, on average, the fastest saccades. Rather, the data is best characterised as follows: IOR is predominantly retinotopic for fast saccadic responses at short post-saccadic delays, and predominantly spatiotopic for slow saccades at long post-saccadic delays (notably 634ms[11]).

The present results, although new within the context of IOR, are entirely consistent with the broader literature on remapping for visual stability. It appears that the same factors that determine the reference frame of IOR also determine the reference frame of

---

11   It is not entirely clear why the difference between spatiotopic and retinotopic IOR for slow saccades has largely vanished at the 1400ms SOA/ 922ms post-saccadic delay (although it is still quantitatively present). Possibly, this is because the general IOR effect is dissipating at this point in time.

many other attentional, visuomotor, and visuotactile phenomena (Azanón, Longo, Soto-Faraco, & Haggard, 2010; Azañón & Soto-Faraco, 2008; Blohm et al., 2005; Golomb et al., 2008, 2011; Overvliet et al., 2011). Most directly relevant here are the studies by Golomb and colleagues (2008, 2011), in which it was shown that attentional facilitation is predominantly retinotopic immediately following an eye movement, and predominantly spatiotopic at longer post-saccadic delay. This clearly resembles the effect of post-saccadic delay that we observed for IOR (Mathôt & Theeuwes, 2010b). Furthermore, the effect of saccade latency is consistent with a study by Blohm and colleagues (2005) on memory guided saccades. They found that fast memory guided saccades are predominantly directed at the retinotopic location of a memorised target, whereas slow saccades are predominantly directed at the actual, spatiotopic location of a target. Remarkably, a study by Overvliet and colleagues (2011) suggests that these effects extend even to visuotactile remapping (i.e., the remapping from an anatomical to a world-centred frame of reference). In their study, participants made a saccadic eye movement to a hand that received tactile stimulation, while (in the crucial condition) their hands were crossed. Their findings parallel the results reported here: The eyes tended to go to the incorrect (anatomical) side when the delay between the tactile stimulation and the saccade cue was short (see also Azanón et al., 2010; Azañón & Soto-Faraco, 2008) and when the saccadic response was fast. We therefore believe that the present results should not be understood solely in the context of IOR, but in the broader context of the coordinate transformations that, among other things, underlie visual stability (Mathôt & Theeuwes, 2011a).

From a neurophysiological point of view, we believe that the distinct pathways hypothesis favoured by Blohm and colleagues (2005) offers a plausible framework to accommodate the present results, particularly the effect of saccade latency. They suggest that very fast saccades are driven by a direct sub-cortical pathway that relies on a 'raw' representation of the environment (Krauzlis & Stone, 1999; Munoz & Wurtz, 1992). This representation is retinotopic for visual input (Blohm et al., 2005), anatomical for tactile input (Azañón & Soto-Faraco, 2008; Overvliet et al., 2011), and more generally reflects the native frame of reference of the input modality. In contrast, slow saccades are mediated by the same parietal and frontal cortical areas that have been implicated in remapping and thus rely on spatially updated representations (Duhamel et al., 1992; Morris et al., 2012; Sapir et al., 2004; van Koningsbruggen et al., 2010). According to this view, the extent to which IOR is retinotopic or spatiotopic depends, at least in part, on the extent to which direct and indirect pathways are involved in the generation of the response.

In summary, we have shown that oculomotor IOR is predominantly retinotopic (gaze-centred) immediately following an eye movement, but particularly for fast saccadic responses, and predominantly spatiotopic (world-centred) at longer intervals after an eye movement, but particularly for slow saccadic responses. This complements our previous report (Mathôt & Theeuwes, 2010b) by showing that the reference frame of IOR depends not only on the time at which the saccade target is presented relative to a preceding eye movement (post-saccadic delay), but also on the speed of the saccadic response (saccade latency). We have stressed that our claim should not be taken to mean that early spatiotopic effects do not exist, but that there is a pronounced difference in the time-course of retinotopic and spatiotopic IOR. Furthermore, the effect of saccade latency on the reference frame of IOR potentially reconciles seemingly conflicting reports (Hilchey et al., in press; Mathôt & Theeuwes, 2010b; Pertzov et al., 2010): Differences in paradigm and trial exclusion criteria may have resulted in different distributions of saccade latencies, in turn affecting the relative strength of retinotopic and spatiotopic IOR. We have suggested that different pathways may be involved in fast and slow saccadic eye movements, relying respectively on retinotopic and remapped (behaviourally spatiotopic) representations (Blohm et al., 2005; Krauzlis & Stone, 1999; Munoz & Wurtz, 1992). Finally, we have argued that the present results are likely not limited to IOR, but extend to many attentional, visuomotor, and visuotactile phenomena.

∞

V. ACTION, PERCEPTION, AND SMOOTH PURSUIT INDUCED MISLOCALISATION

Abstract — We show that a stimulus that is briefly presented during a smooth pursuit eye movement is mislocalised in the direction of pursuit. We argue that this reflects a retinotopic bias. When participants localise the stimulus by making a saccadic eye movement towards its (remembered) location, the retinotopic bias is transient: It occurs only for very fast saccades, but at a larger temporal scale visual stability is preserved (localisation is almost veridical). However, when participants report the location of the stimulus indirectly, by comparing it to a subsequently presented probe stimulus, the retinotopic bias is persistent across a wide range of retention intervals. We argue that this supports a dissociation between visual stability for action-related tasks and perceptual tasks. More specifically, visual stability is preserved only when you engage directly with a stimulus' location, such as when picking up, making an eye movement towards, or attending to a stimulus.

∞

In order to interact effectively with our environment, it is crucial that we do not confuse our own movements with movements of the objects around us. For example, as I am typing this there is a cup of coffee on my desk. When I reach for this cup, my gaze will shift momentarily to the cup's handle. But not for very long. My eyes will quickly shift elsewhere, to my hand, perhaps, or back to the computer monitor. And with each eye movement the position of the cup on my retina changes. Yet my reaching movement, which is ultimately guided by the retinal position of the cup, is unperturbed. And introspectively I do not feel that the cup is displaced with every eye movement that I make. In fact, I am hardly aware that I make eye movements at all. Apparently, therefore, I process visual input in such a way that my behaviour and subjective experience is affected very little by the continuous retinal shifts that are characteristic of visual input at the lowest level (Mathôt & Theeuwes, 2011a; Tatler & Land, 2011).

Demonstrations of a persistent retinotopic bias

Yet many experiments have shown that our subjective sense of visual stability is deceiving, and that we are often unable to distinguish self-generated movement from real movement 'out there'. For example, when you track a flying bird with your eyes (a type of eye movement called 'smooth pursuit'), static objects in the background will subtly appear to move in the direction opposite from your smooth pursuit eye movement. This illusion, first described somewhat anecdotally by Filehne (1922; see also Mack & Herman, 1973), illustrates that we systematically misattribute self-generated retinal motion to static objects. The size of this effect is about 20%, so that 10°/s smooth pursuit will induce a 2°/s illusory motion of static objects (e.g., Thier, Haarmeier, Chakraborty, Lindner, & Tikhonov, 2001). A related phenomenon occurs when people indicate the position of a stimulus that is briefly presented during a smooth pursuit eye movement. The typical finding, dating back to experiments by Hazelhoff and Wiersma (1924), is that stimuli are mislocalised in the direction of smooth pursuit (Brenner, Smeets, & van den Berg, 2001; Mateeff et al., 1981; Mitrani & Dimitrov, 1982; Mateeff et al., 1982; Matsumiya & Uchikawa, 2000). This pattern of smooth pursuit-induced mislocalisation is analogous to the Filehne illusion, in the sense that perception of an object is not veridical (i.e. does not match reality), but driven partly by the location of the object on the retina. In other words, visual perception suffers from a persistent retinotopic bias. An even more extreme example of persistent retinotopy, discussed in Chapter vi of this thesis, comes from studies on the reference frame of the tilt-adaptation aftereffect. Tilt-adaptation is the phenomenon that, after prolonged exposure to a grating of a particular orientation (the adapter), a subsequent grating (the tester) is perceived as tilted 'away' from the adapter orientation (Gibson & Radner, 1937). Crucially, this effect occurs only if the adapter and the tester are presented at the same retinal location (for similar demonstrations, see Afraz & Cavanagh, 2009; Knapen et al., 2009, 2010; Wenderoth & Wiese, 2008; but see Ezzati et al., 2008; Melcher, 2005, 2008a, 2008b; Wittenberg et al., 2008).

Demonstrations of a transient retinotopic bias

The studies described above indicate that, contrary to subjective experience, perception is not fully decoupled from gaze position. It *does matter* where stimuli fall on the retina. This is surprising for a number of reasons. Firstly, persistent retinotopy is not evident during our daily life interactions with objects, as a simple 'experiment' will illustrate: Track the index finger from your left hand with your eyes, while you move it with a moderate speed from left to right. At the same time, pick up an object with your right hand. Although

you will find it difficult not to look at your right hand, and the exercise will likely feel somewhat awkward, you will be able to pick up the object without too much trouble. And, certainly, your (visually guided) reaching movement will not be influenced by the self-generated retinal motion in any obvious way, such as a pronounced misreach in the direction of your smooth pursuit eye movement. This suggests, in apparent conflict with the studies reviewed above, that humans possess a remarkable degree of spatial constancy across eye movements. Phrased differently, it appears that, in some cases at least, it *does not (or hardly) matter* where stimuli fall on the retina. A number of recent studies have investigated this more systematically, and have converged on the same conclusion. For example, in a noteworthy series of studies by Golomb and colleagues (2008, 2011; also see Mathôt & Theeuwes, 2010a, 2010b), participants were required to attend to a location, and subsequently make a saccadic eye movement to another location. After the eye movement, the locus of attention was determined by presenting a probe stimulus at various locations across the visual field. Response times were used as a measure of attention (i.e. participants respond faster to stimuli presented at an attended location, cf. Figure iii.2). Their crucial finding, which has proven consistent across a range of paradigms, was that attention initially moves along with the saccadic eye movement (but see Pertzov et al., 2010). However, this retinotopic trace lasted only very briefly, and at longer delays after the eye movement attention once again resided at the originally attended (spatiotopic) location. In other words, there was some evidence of a transient retinotopic bias, but it was not persistent: After the eye movement, visual stability was momentarily lost, but it was quickly restored.

This pattern of results (a transient retinotopic trace followed by a quick recovery of visual stability) is often taken to reflect an active process that maintains accurate positional information across eye movements. This process is called 'remapping' or 'spatial updating'. There is little consensus about the details and neural mechanisms of remapping, but the common notion is that it operates during a brief temporal window surrounding each eye movement (for various perspectives, see Cavanagh et al., 2010; Hamker et al., 2008; Mathôt & Theeuwes, 2011a; Morris et al., 2012; Wurtz, 2008). As Golomb and colleagues (2008) have shown, during this 'window of instability' the retinotopic nature of the visual system can be exposed. But at a larger temporal scale, remapping ensures that visual stability is nearly perfect.

Taken together, experiments have shown that there is no singular answer to the question of whether, and to what extent, visual stability is preserved across eye movements. Some visual phenomena, such as adaptation aftereffects, are entirely retinotopic (Chapter vi of

this thesis; Afraz & Cavanagh, 2009; Knapen et al., 2009, 2010). Other experimental paradigms, notably those that place a high demand on visual attention and eye movements, reveal that perception is effectively decoupled from gaze position, with the exception of a brief 'window of instability' around each eye movement (Golomb et al., 2008, 2011; Mathôt & Theeuwes, 2010a, 2010b; Rolfs et al., 2011). Still other studies show an intermediate pattern of results: A persistent, but partial dependence on gaze position (Brenner et al., 2001; Filehne, 1922; Mack & Herman, 1973; Mateeff et al., 1981; Thier et al., 2001). The obvious question is therefore: What determines whether visual stability is preserved? Is there a pattern that can explain why some studies find persistent retinotopy, whereas other studies find a transient retinotopic trace followed by a quick recovery of visual stability (i.e. remapping)?

Visual stability in action-related and perceptual tasks

In an attempt to reconcile the many seemingly contradictory results, Bays and Hussain (2007) proposed that visual stability is preserved in 'action-related' tasks, but not in 'perceptual' tasks. Their proposal parallels the classic notion that perceptual information, such as features (colour, etc.) and identity (and possibly conscious visual perception; Milner, 1998), is processed in ventral visual brain areas, whereas action-related information, such as position and shape, is processed in dorsal visual brain areas (Goodale & Milner, 1992; Mishkin, Ungerleider, & Macko, 1983). The distinction between action-related and perceptual tasks is well supported, but nevertheless fuzzy and a potential source of confusion. Therefore, we propose the following guideline as a useful, albeit imperfect means of differentiation: Experimental tasks that require a direct response towards or upon some target stimulus are 'action-related'. Examples of this are studies in which participants make an eye or hand movement to a target stimulus (e.g., Cameron, Enns, Franks, & Chua, 2009; Mathôt & Theeuwes, 2010b). And, in line with the pre-motor theory of attention, which posits that attending to a location is equivalent to programming an action to that location (Craighero & Rizzolatti, 2005; Rizzolatti et al., 1987), attentional/ cuing paradigms also belong to this category (e.g., Golomb et al., 2008; Mathôt & Theeuwes, 2010a, 2010b; Rolfs et al., 2011): Attentional/ cuing paradigms have a 'covert action' component. In contrast, experimental tasks that require an indirect response, conceptually detached from the location of the stimulus, and do not explicitly manipulate the focus of attention are 'perceptual'. Obvious examples of this are adaptation studies in which participants judge the orientation (or colour, gender, etc.) of a target stimulus (e.g., Knapen et al., 2009, 2010; Wenderoth & Wiese, 2008). But, perhaps counter-intuitively, by our criterion many localisation studies are also perceptual, because participants generally

report the location of a target verbally or by indicating it on a virtual ruler, rather than by engaging directly with the target location (e.g., Kaminiarz, Krekelberg, & Bremmer, 2007; Mateeff et al., 1981).

A pattern emerges if we use this guideline to classify studies as either action-related or perceptual. In line with Bayes and Husain's (2007) proposal, those studies that have demonstrated visual stability across eye movements have all used attentional/ action-related paradigms (e.g., Blohm et al., 2005; Golomb et al., 2008, 2011; Mathôt & Theeuwes, 2010a, 2010b). With few exceptions (but see Ezzati et al., 2008; Melcher, 2005, 2007; Melcher & Fracasso, 2012; Melcher & Morrone, 2003; Wittenberg et al., 2008), all other studies have revealed persistent retinotopy, corresponding to a lack of visual stability (e.g., Kaminiarz et al., 2007; Knapen et al., 2009, 2010; Mateeff et al., 1981; Wenderoth & Wiese, 2008). Furthermore, brain areas that are believed to be involved in maintenance of visual stability are located in the dorsal stream of visual processing (Wurtz, 2008). Notably, the lateral intraparietal area (LIP) has been implicated in remapping for visual stability, either through neurons that update receptive fields in anticipation of saccadic eye movements (i.e., predictive remapping; Duhamel et al., 1992; Kusunoki & Goldberg, 2003), or through an eye position signal that could be integrated with retinotopically organised visual input (Morris et al., 2012). Attempts to demonstrate similar properties in ventral brain areas, such as MT, have failed (Hartmann, Bremmer, Albright, & Krekelberg, 2011; Ong & Bisley, 2011). In addition, transcranial magnetic stimulation (TMS) of dorsal areas disrupts visual stability across saccades (Morris, Chambers, & Mattingley, 2007; van Koningsbruggen et al., 2010), whereas, to the best of our knowledge, there are no reports of similar disruptions following TMS of ventral areas. A final consideration in support of Bays and Hussain's (2007) proposal is an ecological one. If we assume that remapping for visual stability requires resources, and that the brain operates efficiently, we would expect visual stability to be maintained only when this serves a clear purpose. And, contrived experimental paradigms notwithstanding, it is difficult to think of situations where visual stability would be useful, other than those that involve a direct interaction with an object. Phrased differently, accurate positional information is important only when you want to reach out and touch an object (or attend to it, make an eye movement to it, etc.).

The present study

The proposal that remapping for visual stability occurs only when there is a direct link between perception and action is attractive (cf. Bays & Husain, 2007). And, by and large, it seems to be supported by the evidence. But it is difficult to derive strong conclusions from the available evidence, because it involves a direct comparison between studies that have

used very different measures and methodologies. Therefore, in the present study we investigate visual stability for action and perception separately, but in comparable experimental paradigms (cf. Dyde & Milner, 2002). To this end, we designed a paradigm in which participants tracked a moving dot with their eyes. In addition, participants localised a target stimulus that was presented very briefly during the smooth pursuit eye movement. In Experiment 1, participants localised the target by making a saccadic eye movement to its (remembered) location (cf. Blohm et al., 2005; McKenzie & Lisberger, 1986; Schlag, Schlag-Rey, & Dassonville, 1990). Following the criterion outlined above, this is an action-related task, because participants engage directly with the target location. In Experiments 2A and 2B, participants localised the target by reporting whether a subsequent probe stimulus was presented to the left or to the right of the target. Following our criterion, this is a perceptual task, because the response (a left/ right button press) does not map directly onto the location of the target.

Based on our prior studies and the broader literature reviewed above, we expect to find a degree of retinotopy in all cases. Yet we expect the action-related and perceptual tasks to differ in the time course of the retinotopic bias. More specifically, in the action-related task (Experiment 1) we expect to find a transient retinotopic component, followed by a gradual, but fast recovery of visual stability. In contrast, in the perceptual task (Experiments 2A and 2B) we expect to find persistent retinotopy. To map out the time course of the retinotopic bias, we rely on the assumption that visual stability during smooth pursuit relies on a remapping process that compensates for eye movements, but always lags behind the eye movement by a certain fixed interval. This is clearly an oversimplification, which notably disregards the predictive component of remapping (Duhamel et al., 1992; Hunt & Cavanagh, 2011; Mathôt & Theeuwes, 2010a; Nakamura & Colby, 2002; Rolfs et al., 2011). However, it offers a convenient working model to construct our hypotheses, and a wide range of relevant findings can be accommodated by this simplified model of remapping: It takes some time for visual stability to be restored after a saccadic eye movement (Golomb et al., 2008; Mathôt & Theeuwes, 2010b; Morris et al., 2012); Fast saccadic eye movements show a stronger retinotopic bias than slow saccadic eye movements (Blohm et al., 2005); Memory guided saccades are aimed at the retinotopic location of a target if the retention interval is (very) brief (Gellman & Fletcher, 1992; McKenzie & Lisberger, 1986), but at the target's actual, spatiotopic location if the retention interval is long (Ohtsuka, 1994; Schlag et al., 1990). In the General discussion, we will come back to this, and consider more detailed models of remapping for visual stability in light of our results.

In Experiment 1, we exploit the inherent variability of the participants' saccade latencies (saccadic response times). We predict that very fast responses are retinotopically driven, measured as a systematic mislocalisation of the target in the direction of smooth pursuit. In contrast, slow responses are expected to be more accurate, directed at the actual, spatiotopic location of the target. In Experiments 2A,B, we explicitly manipulate the interval between the presentation of the target and the probe stimulus. Since we do not expect remapping to occur in these experiments, we expect to find a persistent retinotopic bias, even for longer target-to-probe delays (unlike for memory guide saccades, cf. Ohtsuka, 1994; Schlag et al., 1990).

Finally, in all experiments a salient background texture was presented throughout each trial. This background texture either matched the participants' smooth pursuit eye movement, did not move at all, or moved in the opposite direction from the smooth pursuit. This manipulation allowed us to investigate the extent to which participants anchor their localisation judgement to the visual background. Previous studies have shown that, at least in some cases, stimuli are localised relative to a moving visual background (Brenner & Smeets, 1997; Mateeff et al., 1982). However, it is not clear whether this is a robust phenomenon, whether it generalises to the current paradigm, and how we should expect the effect of background information to differ between the action-related and perceptual experiments in the present study.

Experiment 1

The aim of Experiment 1 was twofold. Firstly, we wanted to replicate the previously reported pattern of a transient retinotopic bias, followed by a gradual recovery of visual stability (cf. Blohm et al., 2005). Specifically, we expected that fast saccadic eye movements towards a (remembered) target stimulus are predominantly retinotopic (i.e. directed at the retinal location of the target), whereas slow saccades are predominantly spatiotopic (i.e. directed at the actual target location). Secondly, we wanted to investigate whether target localisation is anchored to the visual background. An important characteristic of Experiment 1 is that the response (a saccadic eye movement towards the target) maps directly onto the perceived location of the target.

Materials and methods

Participants, software, and apparatus

10 observers, including one of the authors (S.M.), participated in the experiment (age range 22-52, 7 female, all right-handed, all reported normal or correct vision). The

experiment was conducted with approval of the local ethics committee of the VU University Amsterdam and was in accordance with the Declaration of Helsinki. Eye position data was recorded using an Eyelink 1000 (SR Research, Mississauga, ON, Canada), a video-based eye tracker sampling at 1000Hz. Stimuli were presented on a 21" CRT display, with a refresh rate of 100Hz. The experiment was created using OpenSesame (Mathôt, Schreij, et al., 2012) and PsychoPy (Peirce, 2007).

Procedure and design

Before every trial, a small, white fixation dot ($r = 0.3°$, $L_v = 87$ cd/m^2) was presented on a uniform grey background ($L_v = 19$ cd/m^2) for 750 ms, located 15° to the left or the right of the display centre (randomized within blocks), followed by the presentation of a white cross ($r = 0.3°$) at the same location. Drift correction was performed automatically when participants fixated on this cross, after which the trial started.

A schematic example trial is shown in Figure v.1a. Every trial started with the presentation of a white fixation dot ($r = 0.6°$) at the location that was used for drift-correction. During an interval of 200 ms, the background texture gradually appeared. The background texture was a medium contrast ($L_v = 11$ cd/m^2 to $L_v = 28$ cd/m^2), medium frequency ($f = 0.20$ cycles/°), sinusoid luminance modulation. After an additional 750 ms, the fixation dot started to move horizontally across the screen with a constant velocity ($v = 9.4°$/s). If the dot was initially on the left side of the display, it moved to the right, and vice versa. The background texture either matched the movement of the fixation dot (same direction condition), moved with the same speed in the opposite direction (opposite direction condition), or did not move at all (no background motion condition; randomized within blocks; see Figure v.1c). Participants were instructed to track the moving the moving dot with their eyes.

After a random interval (940 ms – 2280 ms), a target (identical to the moving dot) was presented for 100 ms, 9.4° above or below the moving dot (randomized within blocks). The horizontal position of the target was randomly jittered (sampled from a uniform distribution) around the horizontal position of the moving dot, with a maximum displacement of 3.1° to either side. Participants were instructed to make an eye movement to the location of the target as soon as it appeared. The instruction emphasised both accuracy and speed. 300 ms after an eye movement to the target had been detected, the background texture faded out during 200 ms, after which the next trial started. The experiment consisted of 240 trials in 10 blocks, preceded by 24 practice trials.
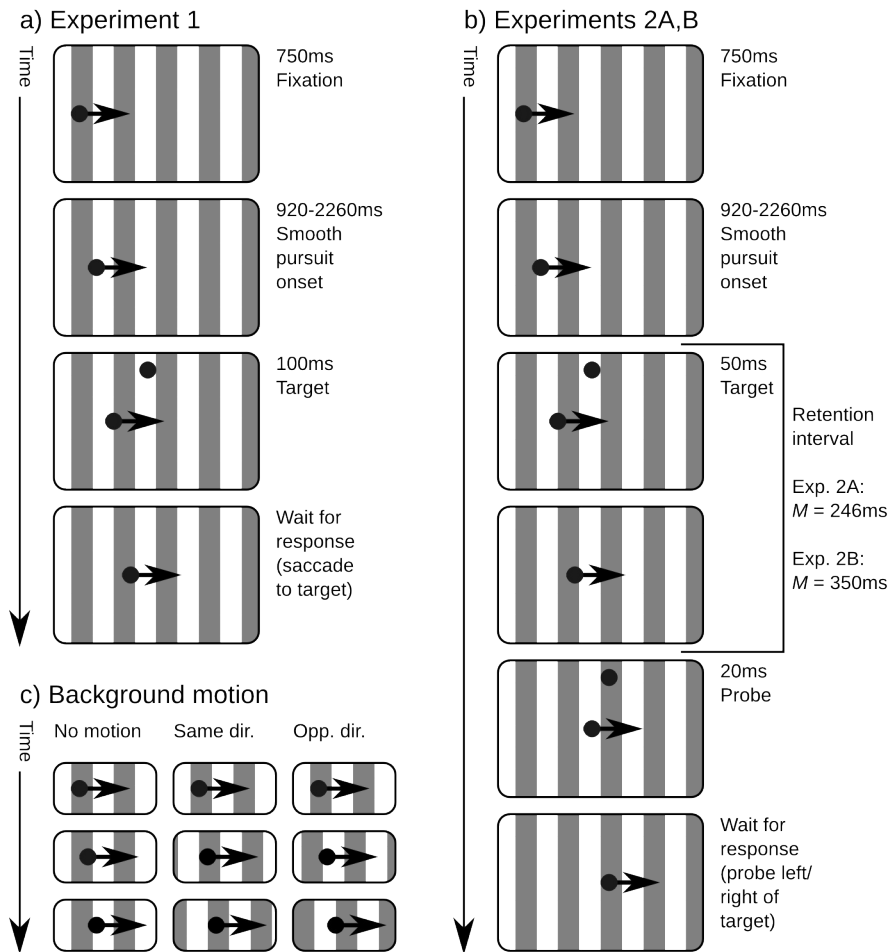
Figure v.1. a) Schematic example trial of Experiment 1. (This is a print-optimized schematic, see text for details.) Participants tracked a horizontally moving dot with their eyes, until a target appeared. Participants responded by making a saccadic eye movement towards the (remembered) target location. b) Schematic example trial of Experiments 2A,B. Participants tracked a horizontally moving dot throughout the trial. Participants responded by indicating, with a button press, whether the probe stimulus was presented to the left or to the right of the target. c) In the no background motion condition, the background was static. In the same direction condition, the motion of the background matched that of the smoothly moving dot. In the opposite direction condition, the background moved in the opposite direction from, but with the same speed as the smoothly moving dot.
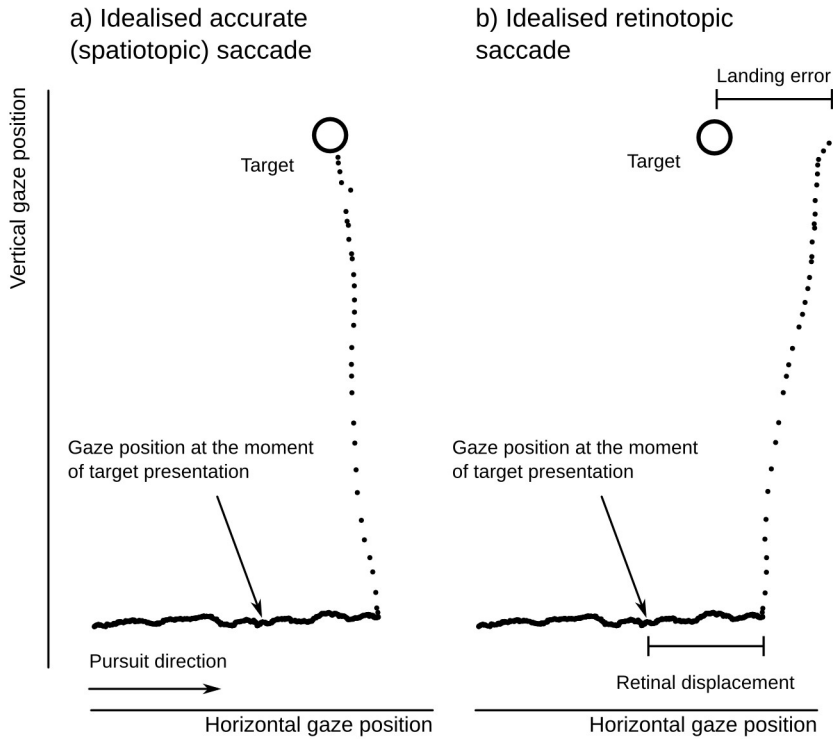
Figure v.2. a) An idealised example of an accurate, spatiotopic response (not actual data). Despite the retinal displacement that has occurred after the target has been presented, the saccadic response is directed accurately at the (remembered) target location. b) An idealised example of a retinotopic response (not actual data). The saccadic response overshoots the (remembered) target location by the amount of the retinal displacement that has occurred after the target has been presented. Effectively, therefore, the saccade has been directed at the retinal location of the target.

Analysis and results

Description of dependent and pseudo-independent measures

Participants did not respond instantaneously when the target was presented. Up until the moment that the eyes set in motion ($M = 259$ms, $SE = 10$), participants continued to track the smoothly moving dot, more or less unperturbed (although the eyes gradually slow down to about 6.5°/s, see Figure v.4). We used *saccadic latency*, the time it took for the eyes to set in motion after the presentation of the target as our first (pseudo-)independent variable. Saccadic latencies were divided into 5 bins (quintiles), determined separately per

subject and background motion condition. Our second independent variable was the *background motion condition*.

Our measure of interest was the *landing error*, the extent to which saccades overshoot the target location (see Figure v.2b). There are a number of ways in which this can be quantified. One way is to use a compensation index (CI), which is a measure of how much participants have compensated for the retinal displacement that has occurred between the moment of target presentation and the onset of the saccadic eye movement (cf. Blohm et al., 2005). However, although it is an elegant measure, the CI can be misleading. Specifically, if participants always overshoot the target by a fixed amount, the CI will increase with increasing retinal displacement. This would lead one to conclude that participants become less retinotopically driven with increasing retinal displacement, even though in this case it would be more informative to speak of a fixed bias. For this reason, we opted to use the landing error directly as dependent measure, and provide the retinal displacement values separately. (Retinal displacement is the result of smooth pursuit during the interval target presentation and saccade onset, and is therefore strongly correlated with saccade latency.) Landing error was defined as the distance between the endpoint of the saccade and the location of the saccade target in the horizontal direction, with positive values reflecting an error in the direction of the smooth pursuit.

Saccade detection, exclusion criteria, and alpha level

Saccades were detected using a velocity/ acceleration threshold ($v > 35°/s$ or $a > 9500°/s/s$). We looked exclusively at the first saccade with a minimum size of 2° after the target had been presented. The following criteria were used to exclude trials from analysis: Saccade latency was less than 100 ms (i.e., the target was still visible at the time of response; 2.8%) or more than 1000 ms (<0.1%); The target was missed by more than 4° in the horizontal plane (1.2%); Gaze deviated more than 4° from the moving dot prior to the presentation of the target (0.3%). In total, 95.6% of the trials were included in the analysis. (We chose liberal inclusion criteria, because all subjects were able to execute the task without any trouble, and we did not want to risk excluding many trials based on eye tracking artefacts.) An alpha level of .05 was used for all statistical tests.

Effect of background motion and saccade latency on landing error

Because background motion had an effect on saccade latency (see the next section), we performed a matching procedure to equate saccade latency across background motion conditions. This allowed us to investigate the direct effect of background motion on landing error, beyond the indirect effect of background motion via saccade latency. On a

Figure v.3. Results of Experiment 1. a) For the fastest saccades, the landing error was about 60% of the retinal displacement. For the slowest saccades this decreased to about 10%. Retinal displacement is indicated next to the symbols. b) The landing error is biased in the direction of the background motion, but overall pattern of results was the same across background motion conditions.

trial by trial basis, we created matched triplets that differed no more than 2ms in saccade latency. Each triplet contained one trial from each background motion condition. 76.7% of all valid trials could be matched in this way. On this subset of trials, a repeated measures Analysis of Variance (ANOVA) was conducted with saccade latency quintile and

background motion condition as within-subject factors, and landing error as dependent variable (Figure v.3). This revealed an effect of saccade latency quintile, $F(4,36) = 5.6$, $p <$ .001, such that the landing error was largest for fast saccades, and an effect of background motion condition, $F(2,18) = 7.3$, $p < .001$. An interaction between saccade latency quintile and background motion condition was not found ($F < 1$). Two-tailed paired samples t-tests revealed that the effect of background motion was such that landing error was larger for the same direction condition, $M = 0.54°$, $SE = 0.11$, relative to the no motion, $M = 0.33°$, $SE = 0.14$, $t(9) = 3.2$, $p < .05$, and opposite direction conditions, $M = 0.26°$, $SE = 0.15$, $t(9) = 4.0$, $p < .01$.

Effect of background motion on saccade latency

A repeated measures ANOVA using background motion condition as within-subject factor and saccade latency as dependent variable revealed an effect, $F(2,18) = 13.6$, $p < .$ 001. Specifically, saccades were slower in the no background motion condition, $M = 265ms$, $SE = 11.3$, relative to the same direction condition, $M = 254ms$, $SE = 10.5$, $t(9) = 5.5$, $p < .001$, and the opposite direction condition, $M = 258ms$, $SE = 9.8$, $t(9) = 2.9$, $p < .$ 05.

Effect of saccade latency and background motion on end pursuit velocity

Although the smooth pursuit eye movement continues right until the onset of the saccadic response, the eyes gradually slow down to about 70% of the original pursuit velocity. A repeated measures Analysis of Variance (ANOVA) was conducted with saccade latency quintile and background motion condition as within-subject factors, and pursuit velocity in the 60ms window prior to saccade onset as dependent variable (see Figure v.4). This revealed an effect of background motion, $F(2,18) = 6.6$, $p < .01$, such that pursuit velocity was highest in the same direction condition, and lowest in the opposite direction condition. There was trend towards an effect of saccade latency, $F(4,36) = 2.1$, $p < .1$. This tentative effect was such that pursuit velocity was lowest for intermediate saccade latencies.

Discussion

In line with our expectations, and consistent with previous findings (Blohm et al., 2005), we found that fast saccadic eye movement towards a (remembered) target location are predominantly directed at the retinal location of the target stimulus, whereas slow saccades are predominantly directed at the target's actual, spatiotopic location. We interpret this as a gradual recovery of visual stability: To maintain accurate positional information, a
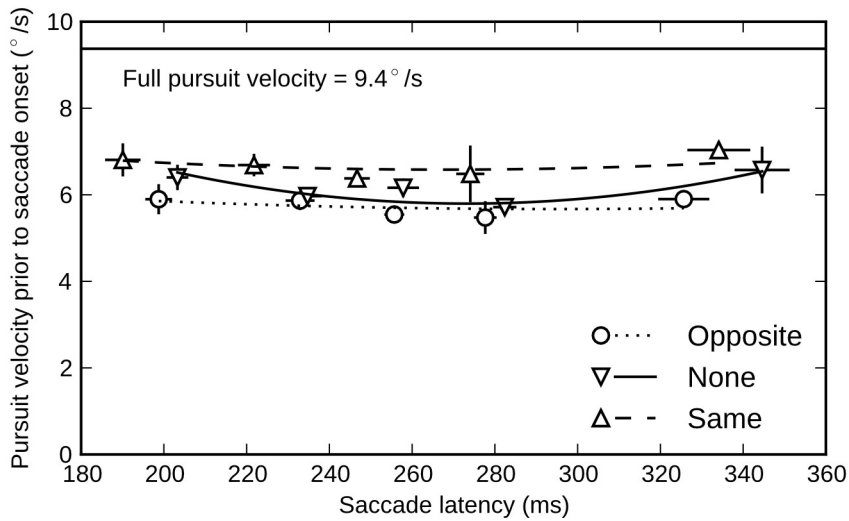
Figure v.4. Effect of background motion condition and saccade latency on smooth pursuit velocity right before the onset of the saccadic eye movement. End pursuit velocity was affected by background motion in a congruent way. Furthermore, there was a trend towards a velocity decrease for saccades with an intermediate latencies. Crucially, however, this tentative effect did not mirror the effect of saccade latency on landing error.

remapping process compensates for the retinal displacements that result from our own eye movements. This process is presumed to take a finite amount of time, so that for very fast saccadic responses visual stability has not yet been restored. We will address this result in more detail in the General discussion.

Furthermore, we found that background information introduces a modest, fixed bias the saccadic localisation judgement. This is particularly true when the background motion matches the smooth pursuit direction, in which case saccadic responses become more retinotopic. However, it is not the case that the background serves as the sole or even primary anchor for the participants' localisation judgements: The overall pattern of gradual recovery is similar across the three background motion conditions.

A final effect, which does not have any direct bearing on our hypotheses, is that saccadic latencies are reduced when the background moves in any direction. The origin of this effect is not entirely clear. A plausible explanation is that background motion, which is subjectively very salient, increases the level of arousal, thus indirectly affecting saccadic latencies. Importantly, we took this effect into account when performing the main analyses.

Experiment 2A

The aim of experiment 2A was to test our prediction that there is a persistent retinotopic bias when participants judge the location of a target stimulus in a way that is perceptual, in the sense that the response does not map directly onto the location of the stimulus. To this end, we designed a paradigm that was very similar to that of Experiment 1, but required a different type of response. Specifically, participants indicated the location of the target stimulus by reporting whether a second stimulus (the probe) was presented to the left or the right of the target. The interval between the presentation of the target and the probe stimulus was matched to the saccade latencies observed in Experiment 1. An important characteristic of Experiment 2A is that the response (a left/ right button press) does not map directly onto the perceived location of the target.

Materials and methods

The methods of Experiment 2A were similar to that of Experiment 1, with the following exceptions. 10 observers, including one of the authors (S.M.), participated in the experiment (age range 20-38, 6 female, all right-handed, all reported normal or correct vision).

The trial procedure was identical to that of Experiment 1 up until the point that the target was presented. To allow for very short retention intervals, the target was presented for only 50ms[12]. At a random time after the target had been extinguished, a probe stimulus (a filled circle, identical to the target) was presented for 20ms. The horizontal position of the probe was randomly jittered (distance sampled from a Gaussian distribution, $\mu = 0°$, $\sigma = 1°$) around the position of the target. The vertical position of the probe was identical to that of the target. On each trial, the target to probe interval, measured as the duration between the onset of both stimuli, was matched (within practical constraints) to the saccade latency on a randomly selected valid trial of Experiment 1 ($M = 246$ms, $SD = 59.2$). Participants reported whether the probe was presented to the right or to the left of the target by pressing respectively the left and right key on a button box. The instruction emphasised accuracy over speed. Participants were instructed to keep their eyes on the horizontally moving dot throughout the trial.

---

12 It has been shown that mislocalisation becomes more pronounced as a stimulus' presentation time decreases, at least for static stimuli (Rotman, Brenner, & Smeets, 2005). Therefore, because we have decreased the duration of the target relative to Exp. 1, we might expect mislocalisation to become more pronounced overall. However, here we are primarily interested in mislocalisation as a function of retention interval, rather than the absolute effect size.

Analysis and results

Description of dependent and independent measures

Our first independent variable was the retention (target-to-probe) interval. This interval was split into 5 bins (quintiles). Our second independent variable was the background motion condition.

Our dependent measure of interest was the perceptual mislocalisation of the target stimulus. This was estimated as follows. First, trials in which the smooth pursuit direction was from right to left were mirrored horizontally. For simplicity, we will discuss the data as though on all trials participants followed the moving dot from left to right. For each subject, using all valid trials, we fitted a sigmoid function with the probe displacement (relative to the target) as predictor, and probability of probe-right response as outcome.

$$PR = \frac{1}{1 + \exp\left(-k * (PD - x0)\right)}$$

Formula v.1

Here, *PR* is the predicted proportion of probe-right responses and *PD* is the probe displacement (positive values reflect a right-wards displacement). *k* and *x0* are sigmoid-specific free parameters that reflect respectively the steepness and horizontal displacement of the curve. Next, we fitted another set of sigmoids, this time separately for each subject and target-to-probe-interval quintile. For these fits, we used the previously determined *k* parameter per subject, leaving *x0* as the only free parameter to be estimated. (Phrased differently, first we determined *k* for each subject. Next we determined *x0* for each combination of subject and target-to-probe-interval quintile.) The reason for keeping *k* constant across target-to-probe interval quintiles, was to boost statistical power by reducing the number of free parameters during the final fit (to obtain *x0*).

For our purpose, *x0* was the crucial free parameter, because it corresponds to the probe displacement at which participants are equally likely to give a probe-right and a probe-left response. *x0* is therefore a measure of the perceptual mislocalisation of the target stimulus.

Exclusion criteria

The following criteria were used to exclude trials from analysis: Participants responded before the presentation of the probe stimulus (<0.1%); Gaze deviated more than 4° from the moving dot prior to response (5.8%). In total, 94.1% of the trials were included in the analysis.

Figure v.5. Results of Experiments 2A,B. a) The perceptual localisation bias increases linearly with retention interval (and thus retinal displacement). Retinal displacement is indicated next to the symbols. b) This relationship is independent of background motion.

Effect of background motion and target to probe interval on perceptual mislocalisation

A repeated measures ANOVA was conducted with target-to-probe interval quintile and background motion condition as within-subject factors, and *x0* (our measure of perceptual mislocalisation) as dependent variable (Figure v.5). This revealed a main effect of target-to-probe interval quintile, $F(4,36) = 18.2$, $p < .0001$, such that the amount of

perceptual mislocalisation increases with increasing target-to-probe interval. No other effects were found (all $F < 1$).

Discussion

The results of Experiment 2A were roughly consistent with our expectations. Indeed, we found persistent retinotopy, which suggests that participants do not remap their representation of the target location to compensate for their own smooth pursuit eye movements.

If we consider the results as a relationship between retinal displacement (which correlates highly with retention interval) and localisation error, we find that mislocalisation is about 50% of the retinal displacement: Participants mislocalise the target by 1° for every 2° of retinal displacement. This indicates that the perceived location of the target is strongly, but not entirely driven by its retinal location. Notably, the size of this effect is much larger than that of the Filehne illusion, in which about 20% of the self-generated retinal motion is erroneously attributed to static objects in the background (Thier et al., 2001).

Furthermore, we found that for very short retention intervals (and thus very small retinal displacements), participants mislocalised the target in the direction opposite from what would be expected based on the retinal displacement. This is, to the best of our knowledge, the first time that a smooth pursuit induced mislocalisation has been demonstrated in the direction opposite from pursuit. The origin of this effect is not entirely clear, but a possible explanation is considered in the General discussion.

Finally, we found no discernible effect of background motion. Participants did not anchor their localisation judgement on the available background information.

Experiment 2B

The aim of Experiment 2B was to investigate whether the linear relationship between mislocalisation and retention interval, as found in Experiment 2A, extends to a wider range of retention intervals. To this end, we increased the target-to-probe interval relative to Experiment 2A.

Materials and methods

The methods were identical to that of Experiment 2A, with the following exceptions. 10 observers, including one of the authors (S.M.), participated in the experiment (age range 18-29, 6 female, 8 right-handed, all reported normal or correct vision). The target-to-probe interval was higher than in Experiment 1, but sampled from an identically shaped distribution ($M = 350$ms, $SD = 58.7$).

Analysis and results

The analysis was identical to that of Experiment 2A, with the following exceptions. Two subjects were excluded from analysis, because they responded at chance level (49% and 54% accuracy). Anecdotally, the poor performance of these two subjects appeared to be due to a very pronounced retinotopic bias, so that their response was predicted almost perfectly by the direction of smooth pursuit. (Debriefing suggested that they had not misunderstood the task.) For the remaining 8 subjects, the following criteria were used to exclude trials from analysis: Participants responded before the presentation of the probe stimulus (0.1%); Gaze deviated more than 4° from the moving dot prior to response (2.9%). In total, 97.0% of the trials were included in the analysis.

A repeated measures ANOVA was conducted with target-to-probe interval quintile and background motion condition as within-subject factors, and *x0* (our measure of perceptual mislocalisation) as dependent variable (see Figure v.5). This revealed a main effect of target-to-probe interval quintile, $F(4,28) = 2.9$, $p < .05$, such that the amount of perceptual mislocalisation increases with increasing target-to-probe interval. No other effects were found (all $F < 1$).

Discussion

The results of Experiment 2B are entirely consistent with those of Experiment 2A. Again, we found a robust and persistent retinotopy. Expressed as a relationship between retinal displacement and mislocalisation, the magnitude of the effect was about 40%, roughly comparable to the 50% that we found in Experiment 2A. Also, as in Experiment 2A, there was no discernible effect of background motion. We will discuss these results in more detail in the General discussion.

General discussion

The aim of present study was to investigate the systematic mislocalisation of a target stimulus that is presented briefly during a smooth pursuit eye movement. In three experiments we investigated the time course of mislocalisation, while varying the way in which the localisation judgement was made. We found that when participants reported the location of the target by making a saccadic eye movement to its (remembered) location there was a pronounced retinotopic bias for fast saccades (Blohm et al., 2005), but very little (systematic) mislocalisation for slow saccades. We interpret this result in terms of visual stability for action (Bays & Husain, 2007): If you engage directly with the location of a perceived stimulus, it is crucial that you do not lose track of its location as a result of self-generated (eye) movements. This is achieved by an active remapping mechanism that

underlies visual stability. But this mechanism is imperfect, which in this case resulted in a transient retinotopic bias for fast eye movements.

However, when participants reported the location of the target indirectly, by pressing a button to indicate the relative position of a subsequently presented probe stimulus, the pattern of results was very different. In this case, we found a persistent retinotopic bias. We interpret this finding in terms of (a lack of) visual stability for perception: If you do not engage directly with a stimulus there is usually little benefit in accurately maintaining its location, and there is therefore no remapping process to compensate for eye movements. As a result, the localisation judgement is heavily biased by the retinal location of the target stimulus.

Therefore, we believe that the present results are consistent with a dissociation between visual stability for action-related and perceptual tasks (using the criterion outlined in the introduction). We will discuss the implications of our results in more detail below.

Stability for action: Direct localisation through saccadic eye movements

First, we will consider a potential trivial explanation for the finding that fast saccades are directed at the retinal location of a (remembered) target stimulus, whereas slow saccades are directed at the actual, spatiotopic location of a target. It has been shown that it is easier to make two eye movements in roughly the same direction, compared to two eye movements in opposite directions, a phenomenon called 'saccadic momentum' (Wang, Satel, Trappenberg, & Klein, 2011). Possibly, the same principle holds for smooth pursuit, so that a saccade will be biased to go in the same direction as a preceding smooth pursuit movement. This could explain the saccadic overshoot that is typically found in experiments such as the present. Furthermore, one might reasonably argue that the velocity of the smooth pursuit movement just before the onset of the saccadic response (end pursuit velocity) correlates with saccade latency, such that end pursuit velocity is lowest for slow saccades. Assuming that saccadic momentum is modulated by end pursuit velocity, the present results could be restated as follows: Slow responses suffer less from saccadic momentum. It is therefore crucial to note that, although there was a trend towards an effect of saccade latency on end pursuit velocity, this effect did not mirror the effect of saccade latency on landing error. Momentum can therefore not explain the current pattern of results.

In line with previous studies (e.g., Blohm et al., 2005; McKenzie & Lisberger, 1986; Schlag et al., 1990), we interpret the present results in terms of visual stability. More specifically, we interpret the gradual decrease in landing error as reflecting remapping for visual stability, or a gradual increase in the participant's ability to take into account his or

her own smooth pursuit eye movements. So far, our use of the term 'remapping' has been agnostic with respect to any specific mechanism. In light of our results, we will now consider three specific mechanisms, not necessarily mutually exclusive, that have been proposed to underlie remapping for visual stability.

A first possibility, which has recently received considerable attention, is that *predictive remapping* is the sole key to visual stability (Duhamel et al., 1992; Mathôt & Theeuwes, 2011a; Nakamura & Colby, 2002; Wurtz, 2008). According to the predictive remapping hypothesis, the visual system anticipates the visual consequences of an eye movement before it is executed, effectively updating its retinotopic representations ahead of the actual retinal displacement (Duhamel et al., 1992; Mathôt & Theeuwes, 2010a; Rolfs et al., 2011). In its strongest form, predictive remapping is a purely anticipatory process that never plays catch up with eye position. Although predictive remapping has been studied almost exclusively in the context of saccadic eye movements, we can extend this principle to smooth pursuit. This would lead to the prediction that any retinotopic bias should be inverse, opposite from the smooth pursuit movement, due to the anticipatory nature of predictive remapping. This is clearly not what we found, since the retinotopic bias was consistently in the direction of the smooth pursuit (Blohm et al., 2005; Brenner et al., 2001; Kaminiarz et al., 2007; Mateeff et al., 1982, 1981). Therefore, although predictive remapping may be a mechanism that plays a role in visual stability (but see Hamker et al., 2008; Morris et al., 2012; Zirnsak et al., 2010), it cannot, by itself, accommodate the present results.

A second possibility is that visual stability emerges by combining retinotopic representations with an *eye position signal*. This is an old idea (Sperry, 1950; von Holst & Mittelstaedt, 1950), but an appropriate eye position signal has only recently been characterised in detail, in dorsal areas of the macaque visual system (Morris et al., 2012). Updating of the eye position signal starts prior to the onset of a saccadic eye movement, reminiscent of predictive remapping, but is not fully complete until sometime after an eye movement. Therefore, unlike predictive remapping, a model of visual stability based on an eye position signal is consistent with psychophysical studies that have demonstrated a transient, post-saccadic disruption of visual stability (e.g., Golomb et al., 2008; Mathôt & Theeuwes, 2010b; Ross, Morrone, Goldberg, & Burr, 2001). The properties of the eye positional signal during smooth pursuit have not been characterised. However, again, we can extend the general principle from saccadic eye movements to smooth pursuit, by conceptualising smooth pursuit as a sequence of very small saccadic eye movements. This would lead us to cautiously predict that during smooth pursuit the eye position signal

consistently lags behind the actual eye position by a certain fixed amount. With respect to the present experiment, this would lead to the prediction that the absolute mislocalisation is relatively constant, independent of the amount of retinal displacement (although the relative mislocalisation will decrease over time). This prediction does not accommodate our finding that the amount of mislocalisation decreases, also in an absolute sense, as the amount of retinal displacement increases.

A third possibility, favoured by Blohm and colleagues (2005), is that the effect of saccade latency reflects the existence of two *distinct pathways* for saccade programming (Krauzlis & Stone, 1999). One pathway, likely striatal-collicular, is fast, direct, and driven solely by the retinal location of a stimulus. This pathway underlies fast eye movements, which are consequently aimed at the retinal location of stimuli. A second, slower pathway, presumably involving the parietal and frontal cortical areas that have been implicated in visual stability (Duhamel et al., 1992; Morris et al., 2012), is driven by an accurate, remapped representation of the stimulus. This pathway underlies slow eye movements, which are consequently more accurate and hardly affected by a retinotopic bias. Of the models considered here, this explanation accounts best for our finding that visual stability is lost for fast eye movements, but almost fully restored for slow saccadic eye movements. Notably, the distinct pathways hypothesis does not invoke remapping in the conventional sense (i.e. coordinate transformations within brain areas), but relies on the differential involvement of different brain areas.

Stability for perception: Indirect localisation through manual responses

In Experiments 2A,B, participants localised the target by indicating with a button press whether a subsequent probe stimulus appeared to the left or the right of the target. This response was perceptual, in the sense that it was indirect and conceptually detached from the location of the target. Rather than mapping out the mislocalisation as a function of saccade latency, as in Experiment 1, we varied the retention (target-to-probe) interval. Although saccade latency and retention interval are very different measures, most theories on remapping nevertheless predict that visual stability is better preserved for both slow saccadic responses, and longer retention intervals. Support for this assumption comes from localisation studies that have shown, using saccadic responses, that there is a strong retinotopic bias for short retention intervals (Gellman & Fletcher, 1992; McKenzie & Lisberger, 1986) and little or no such bias for longer retention intervals (Ohtsuka, 1994; Schlag et al., 1990).

Our results using an indirect perceptual judgement are emphatically very different from localisation studies using saccadic responses (Blohm et al., 2005; Gellman & Fletcher,

1992; McKenzie & Lisberger, 1986; Ohtsuka, 1994). We found that localisation suffered from a persistent retinotopic bias of approximately 50%, across the entire range of retention intervals that we tested. This supports our hypothesis that the time course of the retinotopic bias depends on the way in which a localisation judgement is made. However, there are a number of puzzling aspects to the results, which warrant further discussion.

Notably, for very short retention intervals, participants mislocalised the target in the direction opposite from smooth pursuit. This is, to the best of our knowledge, the first time that an inverse smooth pursuit-induced mislocalisation effect has been demonstrated. However, the cause of this effect is not entirely clear. One possible explanation is that participants attempt to compensate for their smooth pursuit eye movements, but do not have the amount of compensation depend on the actual amount of retinal displacement that needs to be compensated for. This could explain the overcompensation for small retention intervals (i.e. small retinal displacements) and undercompensation for long retention intervals (i.e. large retinal displacements).

Furthermore, although the retinotopic bias was substantial, it was not complete. The participants' localisation judgements were not based solely on the retinal location of the target. Given our hypothesis that remapping for visual stability does not occur in this type of perceptual task, the obvious question is how participants are nevertheless able, however imperfectly, to detach their localisation judgement from the retinal input. One plausible explanation is that participants also rely on landmarks, such as the edges of the screen, to perform their localisation judgement (Deubel, 2004; Deubel et al., 2010). At first glance, this seems inconsistent with the finding that motion of the background texture, arguably the most prominent landmark, has little or no effect on the localisation judgement. However, it has been shown that objects lose their efficacy as landmarks when they are clearly unreliable, such as when they are moving (Gysen et al., 2002). This could explain why the screen edge, rather than the background texture, would be used as a stable landmark in the present experiments.

Role of background information in visual stability

We consistently found that motion of the background texture has little (Experiment 1) or no (Experiments 2A,B) effect on the localisation judgement. This is surprising, given the well known Duncker illusion (Duncker, 1929; related to the effect described by Filehne, 1922), in which background motion, either real or resulting from smooth pursuit, induces an illusory motion of stimuli that are placed on this background. Clearly, one would expect the Duncker illusion to occur in our experiments, in which there was both self-generated and actual background motion. Yet there was no trace of this illusion in our

results. In Experiment 1 there was some effect of background motion, but it was such that participants mislocalised the target in the direction of the movement, which is opposite from what would be expected based on the Duncker illusion.

Quite possibly, this can be explained by the fact that optical illusions selectively affect particular features, while leaving others untouched (for a review, see Smeets et al., 2002). In the case of the Duncker illusion, there is a pronounced illusory motion, but location judgements (Bacon, Gordon, & Schulman, 1982) and pointing movements toward objects (Bridgeman, Kirch, & Sperling, 1981) are not similarly affected. In fact, if there is a perceived shift in position, it is in the direction of the background motion, such that the background serves as an anchor (Brenner & Smeets, 1997). Qualitatively, this is what we found as well: In Experiment 1, the effect of background motion was such that participants mislocalised the stimulus in the direction of the background motion. Quantitatively, however, this effect was much smaller than the substantial effects demonstrated elsewhere (Brenner & Smeets, 1997). Plausibly, this is because in our study the background texture moved continuously, which may have attenuated its effectiveness as a landmark stimulus (Gysen et al., 2002).

Conclusion

In summary, we have shown that a stimulus that is briefly presented during a smooth pursuit eye movement is mislocalised in the direction of pursuit. We have argued that this reflects a retinotopic bias, such that localisation judgements are biased towards the retinal location of a stimulus. Furthermore, we have shown that this retinotopic bias is transient when participants localise the stimulus by making a saccadic eye movement towards its remembered position. In contrast, when participants report the location of the stimulus indirectly, by comparing it to a subsequently presented probe stimulus, the retinotopic bias is persistent. We have interpreted these results as a dissociation between visual stability for action-related tasks, and visual stability for perceptual tasks (cf. Bays & Husain, 2007). More specifically, visual stability is only preserved when you engage directly with a stimulus, such as when reaching for, making an eye movement towards, or attending to a stimulus' location (following the pre-motor theory of attention; Craighero & Rizzolatti, 2005; Rizzolatti et al., 1987). We have argued that visual stability for action relies on a remapping process that restores stability at a larger temporal scale, but leads to momentary glitches in the form of a post-saccadic retinotopic trace (Golomb et al., 2008; Mathôt & Theeuwes, 2010a, 2010b) and a retinotopic bias for fast responses (see also Blohm et al., 2005).

∞

VI. A RE-INVESTIGATION OF THE REFERENCE FRAME OF THE TILT-ADAPTATION AFTEREFFECT

Abstract — Using a novel, speeded response time paradigm, we show that the location-specific component of the tilt-adaptation aftereffect (TAE) is exclusively retinotopic. Neither at brief nor at longer intervals after a saccadic eye movement did we find any location-specific TAE at the originally adapted location, although we found strong TAE at the location that retinotopically matched the originally adapted location. This result supports the notion that there is little to no trans-saccadic integration of low-level visual information, such as orientation.

∞

A central issue in research on visual stability is the extent to which a detailed representation of our visual surroundings is preserved across eye movements (for a review see Mathôt & Theeuwes, 2011a). We can split this issue into two questions. Firstly, do we maintain a representation of our entire visual environment, or only of a subset of objects, presumably those that are in the focus of attention? And secondly, are the representations that underlie trans-saccadic integration rich, in the sense that they contain detailed information about visual features, form, etc. Or are these representations sparse, perhaps barely more than some positional information, or 'attentional pointers' (cf. Cavanagh et al., 2010), to serve visually guide action?

There is broad consensus on the first question: We do not maintain a cognitive representation of our entire visual environment, at least not if we equate representation with visual awareness (O'Regan & Noë, 2001). Rather, as experiments on change detection (O'Regan et al., 2000; Rensink et al., 1997; Simons & Rensink, 2005) and inattentional blindness (Mack, 2003; Mack & Rock, 1998; Rock, Linnett, Grant, & Mack, 1992) have shown, we are only aware of a very limited number of objects at a time. This is true for perception in general, but also applies to trans-saccadic integration: We preserve only a limited number of objects across eye movements (Irwin, 1992b; Prime et al., 2007).

However, there is considerable controversy surrounding the nature of the representations that underlie trans-saccadic integration. What type of information is preserved across eye movements? The majority of studies point towards the counter-intuitive conclusion that very little information about objects is retained across saccades, even about those objects that are in the focus of attention (Henderson et al., 1987; Irwin, 1996; McConkie & Zola, 1979; for a review, see Mathôt & Theeuwes, 2011a). This has been demonstrated elegantly in a classic study by McConkie and Zola (1979), in which participants read words consisting of letters with randomly alternating case (LiKE tHis). The crucial manipulation was that letter case was reshuffled when participants made an eye movement (e.g., from LiKE tHis to LikE ThiS). The surprising finding was that participants frequently failed to notice this change. Since they were reading, we may assume that the participants were paying attention to the words. Yet this seemingly obvious change went unnoticed, even though it occurred to an attended stimulus in foveal vision. This striking finding clearly suggests that there is little or no trans-saccadic integration of detailed object features.

However, another line of research converges on the opposite conclusion, namely that low-level features are preserved, or 'remapped', across saccades as well. The most convincing evidence for remapping of low-level features comes from studies on adaptation aftereffects across saccades (Biber & Ilg, 2011; Ezzati, Golzar, & Afraz, 2008; Melcher, 2005, 2007, 2008; see also Demeyer, De Graef, Wagemans, & Verfaillie, 2010; Melcher & Fracasso, 2012; Melcher & Morrone, 2003; Wittenberg, Bremmer, & Wachtler, 2008; Zimmermann, Burr, & Morrone, 2011). In general terms, an aftereffect is the phenomenon that, after prolonged exposure to an adapter stimulus, people perceive a subsequently presented tester stimulus as being 'pushed away' from the adapter stimulus. For example, in the case of faces, this means that people perceive an androgynous tester face as male, when it is preceded by a female adapter face (face-adaptation aftereffect; Zimmer & Kovács, 2011). In the case of orientation, this means that people perceive a vertical stimulus as being tilted clockwise, when it is preceded by a counterclockwise adapter stimulus (tilt-adaptation aftereffect, TAE; Gibson & Radner, 1937). The same principle holds for a wide variety of stimulus features, such as form, direction, motion, and numerosity.

It has been long recognised that adaptation aftereffects are partly location-specific, such that the effect is largest when the adapter and tester are presented at the same location. However, it was not clear whether such effects are tied to a retinal location (retinotopic) or a location 'out there' in space (spatiotopic). Therefore, Melcher (2005) set out to investigate the extent to which aftereffects are spatiotopic (i.e. decoupled from gaze). He investigated

four types of adaptation aftereffects (contrast-adaptation, tilt-adaptation, form-adaptation, and face-adaptation) using a straight-forward paradigm. First, an adapter stimulus was presented at fixation. After the adapter had been extinguished, participants made a saccadic eye movement. Finally, a tester stimulus was presented. The participants identified the tester stimulus (for example by indicating the gender for the faces). Crucially, the tester was presented either at the same spatial location as the adapter (the spatiotopic condition), or at a control location. The striking finding was that the extent to which a spatiotopically selective aftereffect was observed depended on the complexity of the stimulus: There was no spatiotopic adaptation for contrast, some for tilt, more for form, and almost complete spatiotopic transfer for faces. Melcher (2005) interpreted this result as showing that detailed visual features are preserved across saccades, even relatively low-level visual features, such as orientation, although to a lesser extent (cf. Merriam et al., 2007).

Although the strongest spatiotopic effects were observed for faces (Melcher, 2005), subsequent research focused primarily on the tilt-adaptation aftereffect (TAE), which has been used to investigate the mechanisms that underlie visual stability in general. For example, based in part on TAE experiments it was concluded that visual stability relies on a remapping process that starts before the onset of a saccadic eye movement (Melcher, 2007), and that trans-saccadic integration of visual features occurs primarily for objects that are in the focus of attention (Melcher, 2008b). In other words, the premise that spatiotopic TAE exists has guided experiments and thinking of many researchers—including ourselves (for reviews, see Mathôt & Theeuwes, 2011a; Wurtz, 2008).

Yet recently, a number of authors have reported purely retinotopic adaptation aftereffects, without any corresponding spatiotopic component (Afraz & Cavanagh, 2009; Knapen et al., 2009, 2010; Wenderoth & Wiese, 2008). Because of the impact of experiments on spatiotopic TAE, the debate over whether this effect exists at all is of considerable theoretical significance. Therefore, the aim of the present study was to re-investigate the reference frame of the tilt-adaptation aftereffect (TAE). In our view, there are two troubling factors in most previous research on the reference frame of adaptation aftereffects. Firstly, most paradigms did not include control locations that were matched with respect to eccentricity and position relative to saccade direction. This made it difficult to convincingly dissociate location-specific from generalised effects. To address this problem we used separate, carefully matched control locations for both the spatiotopic and retinotopic locations. Secondly, the nature of adaptation experiments allows for a confounding influence of the observers' expectations: Trial progression is slow, experiments are tedious, and responses are generally not speeded. Therefore, observers

have ample opportunity to contemplate the goals of the experiment and inadvertently adjust their responses accordingly. This may confound results even if observers are naïve, but is particularly problematic when many (Knapen et al., 2010) or even most (Zirnsak et al., 2011) of the observer are also authors. To alleviate this issue, all observers in the crucial experiments (2 and 3) were naïve and without training as psychophysical observers. Furthermore, observers were instructed to respond as quickly as possible, in order to minimize the opportunity for excogitation. Aside from these points, we have used the same overall methodology as used in many of the studies mentioned above: First, a task-irrelevant adapter stimulus was presented. Next, observers made a saccadic eye movement. Finally, a tester stimulus was presented at either the adapter location (spatiotopic), the location that retinotopically matched the adapter location (retinotopic), or one of two control locations. Observers reported the orientation of the tester stimulus, and we measured the effect of the adapter on the perceived orientation of the tester.

Experiment 1

In a typical study of the tilt-adaptation aftereffect (TAE), participants make a non-speeded report of the orientation of a tester stimulus, which is usually a Gabor patch. By systematically testing a range of tester orientations, one can determine the orientation at which the tester subjectively appears to be vertical. Despite its elegance, this slow paced approach is prone to expectation effects. Therefore, in the present experiments we used a simplified speeded response time task to measure TAE.

The aim of Experiment 1 was to verify that we could obtain a robust TAE effect in a speeded response time task. Furthermore, we wanted to determine the optimal orientation for the adapter and tester grating, and the most sensitive dependent measure (i.e. response times or accuracy).

Methods

8 observers, including one of the authors (SM), participated in the experiment. All participants reported normal or corrected visual acuity. Eye movements were recorded using an Eyelink 1000 (SR Research, Mississauga, Canada, ON), a video based eye tracker sampling at 1000Hz. Stimuli were presented on a 22" CRT monitor, with a resolution of 1024x768px and a refresh rate of 100Hz.

A schematic example trial is shown in Figure vi.1a. Before the start of each trial, a central white fixation cross was presented against a grey background. A drift correction procedure was triggered automatically as soon as a stable fixation was detected, except before the first trial of each block, in which a space bar press was required. Next, the trial
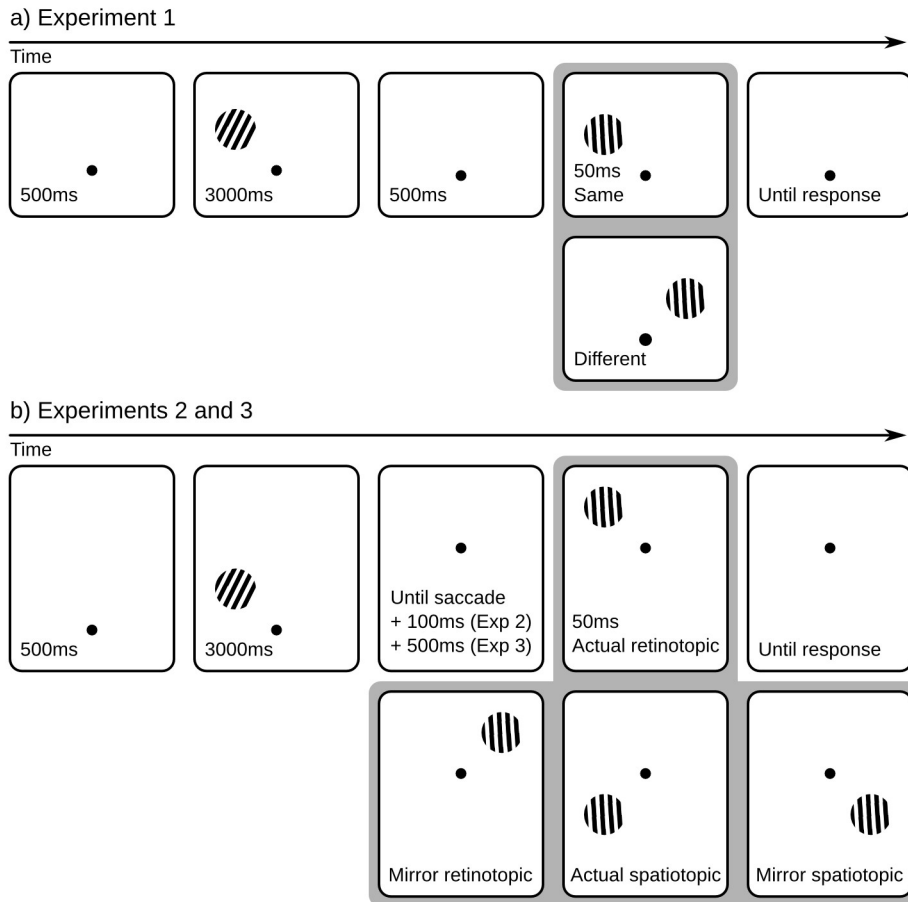
Figure vi.1. a) Schematic example trial for Experiment 1. An adapter stimulus was presented, followed by a tester stimulus. The tester stimulus was presented either at the same location as the adapter, or at a different location. The participant reported the orientation of the tester (rotated left/ right). b) Schematic example trial for Experiments 2 and 3. The tester was presented either at the same location as the adapter (actual spatiotopic), the same retinotopic location as the adapter (actual retinotopic), or one of two control locations (mirror actual and mirror retinotopic).

proper started with the presentation of a central white fixation dot. After 500ms, an adapter stimulus was presented for 3000ms at a fully random location 4.2° from the fixation dot. The adapter stimulus was a sinusoid luminance modulation with a spatial frequency of 2.5 cycles/°, maximum contrast, a linear envelope, and a radius of 4°. The adapter was rotated clockwise or counterclockwise by 30° or 15° (angular) from a vertical orientation. 500ms after the adapter was extinguished, a tester stimulus was shown for 50ms. The tester was presented at the same location as the adapter stimulus, or at 6° distance from the adapter at the same eccentricity from the fixation dot. The tester was rotated clockwise or counterclockwise by 6°, 4°, or 2° (angular). Participants were instructed to report the orientation of the tester stimulus as quickly as possible by pressing the 'z' key on a counterclockwise rotation, and the slash-key on a clockwise rotation.

Tester orientation (2°, 4°, 6°), adapter orientation (15°, 30°), and tester location relative to adapter (same, different) were mixed within blocks. The experiment consisted of 384 trials, divided into 6 blocks, and was preceded by 24 practice trials.

Measure of location-specific tilt-adaptation aftereffect

In a typical tilt-adaptation experiment, the orientation of the tester stimulus is perceived as being tilted slightly away (relative to its actual orientation) from the orientation of the adapter stimulus. In other words, the orientation of the tester will seem more pronounced when it is preceded by an adapter that is oriented in the opposite direction (incongruent; e.g., a 2° tester and a -30° adapter), compared to when it is preceded by an adapter oriented in the same direction (congruent trials; e.g., a -4° tester and -15° orientation). The more pronounced the orientation of the tester appears, the faster and more accurate participants will respond. Therefore, in the current paradigm TAE can be measured as a reverse congruency effect.

We calculated the strength of TAE as the difference in mean response times and accuracy between congruent and incongruent trials. Our measure of interest was the difference in TAE between same location, and different location trials, as this reflects the strength of the location-specific TAE, or LSTAE.

For response times, LSTAE was determined as follows:

$$LSTAE_{rt} = RT_{(same,con)} - RT_{(same,inc)} - RT_{(diff,con)} + RT_{(diff,inc)}$$

Formula vi.2

Here *RT* is the mean correct response time, *same* is the same location condition, *diff* is the different location condition, *con* are congruent trials, and *inc* are incongruent trials. For accuracy, LSTAE was determined as follows:

$$LSTAE_{acc} = Acc_{(same,inc)} - Acc_{(same,con)} - Acc_{(diff,inc)} + Acc_{(diff,con)}$$
Formula vi.3

Here, *Acc* is the proportion of correct trials.

Results

Trials were excluded based on the following criteria: Gaze deviated from than 2° from the fixation dot (7.2%); Response times were below 50ms (-) or 1500ms (0.3%). In total, 92.6% of all trials were included in the analysis. An alpha level of .05 is used throughout the analyses.

A repeated measures Analysis of Variance (ANOVA) was conducted with tester orientation and adapter orientation as within-subject factors, and $LSTAE_{rt}$ as dependent variable. This revealed no effects (Figure vi.2b). A similar analysis was conducted with $LSTAE_{acc}$ as dependent variable. This revealed an effect of tester orientation, $F(2,14) = 5.4$, $p < .05$, such that $LSTAE_{acc}$ was larger for small tester orientations. Tentatively, there was trend towards an effect of adapter orientation, $F(1,7) = 4.8$, $p < .1$, such that $LSTAE_{acc}$ was largest for the 30° adapter orientation (Figure vi.2a).

Finally, we checked whether there was a robust location-specific TAE for the optimal combination of dependent variable ($LSTAE_{acc}$), tester orientation (2°), and adapter orientation (30°). Assuming a uniform distribution, a lower bound of 0 (no difference), and an upper bound of 50 (maximum difference), we determined the Bayes factor: $Bf = 74.8$, $M = 29$, $SE = 10$. Following Jeffreys (1961; reproduced in Wetzels et al., 2011), this constitutes "very strong evidence" for the existence of an effect (H_a).

Discussion

It proved easy to elicit a substantial, location-specific TAE using a speeded response time task. The effect was found, at least qualitatively, across all tester and adapter orientations, and for both response times and accuracy. However, the most robust results were obtained using accuracy as dependent measure, a tester orientation of 2°, and an adapter orientation of 30°. We therefore used these parameters, and this dependent measure for Experiments 2 and 3.

Experiment 2

The aim of experiment 2 was to determine the reference frame of TAE. To this end, we designed a paradigm in which participants made an eye movement between the
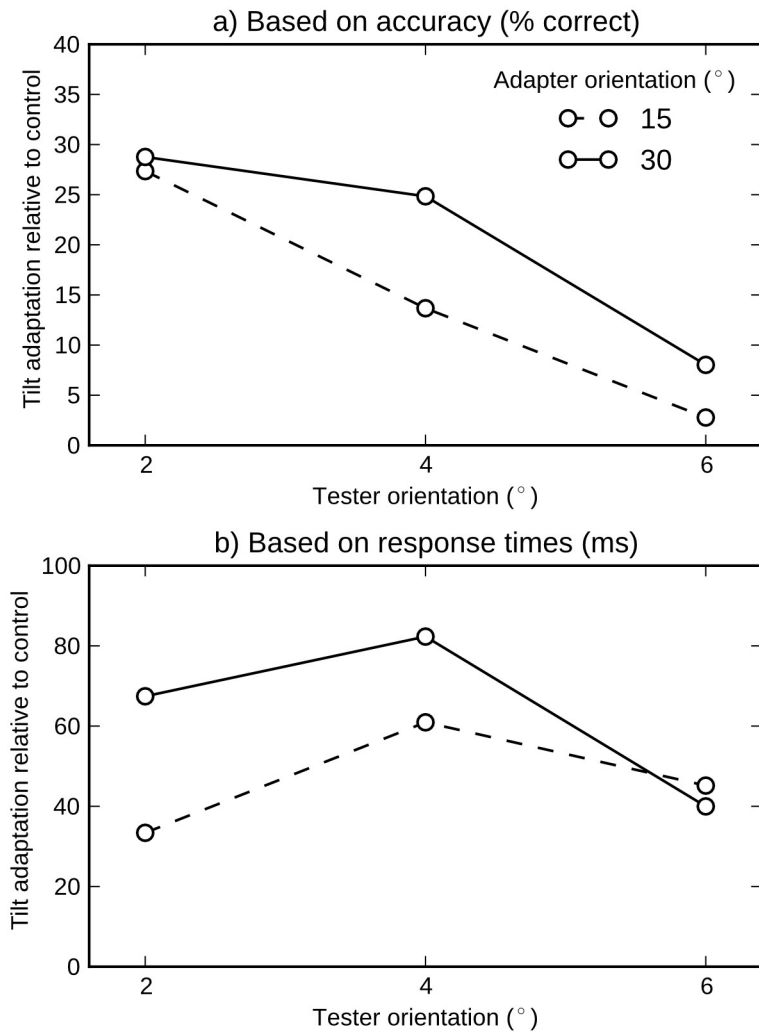
Figure vi.2. a, b) Location-specific TAE (LSTAE) as a function of adapter orientation and tester orientation. a) LSTAE as measured using accuracy. b) LSTAE as measured using mean response times.

presentation of the adapter and the tester stimulus. This allowed us to dissociate the retinotopic (gaze-centred) and spatiotopic (world-centred) component of TAE.

Method

The method was similar to that of Experiment 1, with the following exceptions. 8 new observers participated in the experiment. All were naïve as to the purpose of the study and none were trained psychophysical observers.

A schematic example trial is shown in Figure vi.1b. After the presentation of the adapter stimulus, the fixation dot was displaced 6° to a fully random location, with the constraint that the selected location was always 6.4° away from the display edge. Participants were instructed to match the displacement with their eyes. 100ms after a saccadic eye movement had been detected, the tester stimulus was presented for 50ms. There were four possible stimulus configurations. In the *actual spatiotopic* condition, the tester was presented at the same location as the adapter. In the *mirror spatiotopic* condition, which served as a control for the actual spatiotopic condition, the tester was presented at the location that mirrored the adapter location in the trajectory of the eye movement. In the *actual retinotopic* condition, the tester was presented at the same retinal location as the adapter stimulus. In the *mirror retinotopic* condition, which served as a control for the actual retinotopic condition, the tester was presented at the location that mirrored the retinal adapter location in the trajectory of the eye movement. The final fixation location was used as the initial fixation location for the next trial, so that the paradigm had the appearance of a random walk across the display.

The tester grating was always tilted 2° clockwise or counterclockwise. The adapter grating was always tilted 30° clockwise or counterclockwise. Condition (spatiotopic, retinotopic) and location (actual, control) were randomly mixed within blocks. The experiment consisted of 384 trials, divided into 6 blocks, and was preceded by 24 practice trials.

Measure of tilt-adaptation aftereffect

We used accuracy as our primary measure, although we analysed response times as well. For this experiment we determined TAE directly, rather than the location-specific TAE, to get a more transparent picture of the results.

For accuracy, TAE was determined as follows:

$$TAE_{acc} = Acc_{inc} - Acc_{con}$$

Formula vi.4

For response times, TAE was determined as follows:

$$TAE_{rt} = RT_{con} - RT_{inc}$$
Formula vi.5

Results

Trials were excluded based on the following criteria: The eyes deviated more than 2° from the expected location during the trial (10.3%); Saccade latencies were below 50ms (3.8%) or above 500ms (1.0%); Response times were below 50ms (<0.1%) or above 1500ms (1.1%). In total, 83.7% of all trials were included in the analysis.

A repeated measures ANOVA was conducted with condition (spatiotopic, retinotopic) and location (actual, mirror) as within-subject factors and TAE_{acc} as dependent variable (see Figure vi.3a). This revealed an effect of condition, $F(1,7) = 35.4$, $p < .001$, such that TAE_{acc} was higher in the retinotopic than in the spatiotopic condition, and an effect of location, $F(1,7) = 21.8$, $p < .01$, such that TAE_{acc} was higher at the actual than at the mirror locations. Crucially, there was a condition by location interaction, $F(1,7) = 14.6$, $p < .01$, reflecting that the main effects are driven by a difference between the actual and mirror retinotopic conditions.

A similar analysis with TAE_{rt} as dependent variable, revealed only a main effect of condition, $F(1,7) = 8.0$, $p < .05$. The results from TAE_{rt} qualitatively matched those from TAE_{acc}.

To verify the existence or non-existence of location-specific TAE more rigorously for both the spatiotopic and retinotopic condition, we determined the Bayes factor for the difference between actual and mirror trials in both the retinotopic and spatiotopic condition. We assumed a uniform distribution with realistic lower and upper bounds for the expected difference. We set the lower bound to 0, since this reflects an absolute lack of location-specific TAE. We set the upper bound to 50, since this reflects the largest possible location-specific TAE. For the spatiotopic condition, this gave us the following: $M = 1.67$, $SE = 3.99$, $Bf = 0.14$. Following Jeffreys (Jeffreys, 1961; reproduced in Wetzels et al., 2011), this indicates "substantial evidence" *against* location-specific TAE. For the retinotopic condition, this gave us the following: $M = 29.96$, $SE = 5.87$, $Bf > 1x10^5$. This indicates "decisive evidence" *in favour of* a location-specific TAE.

Discussion

The results show unambiguously that the location-specific aspect of TAE is retinotopic shortly after an eye movement: Tilt-adaptation is anchored to the retina (retinotopic), and
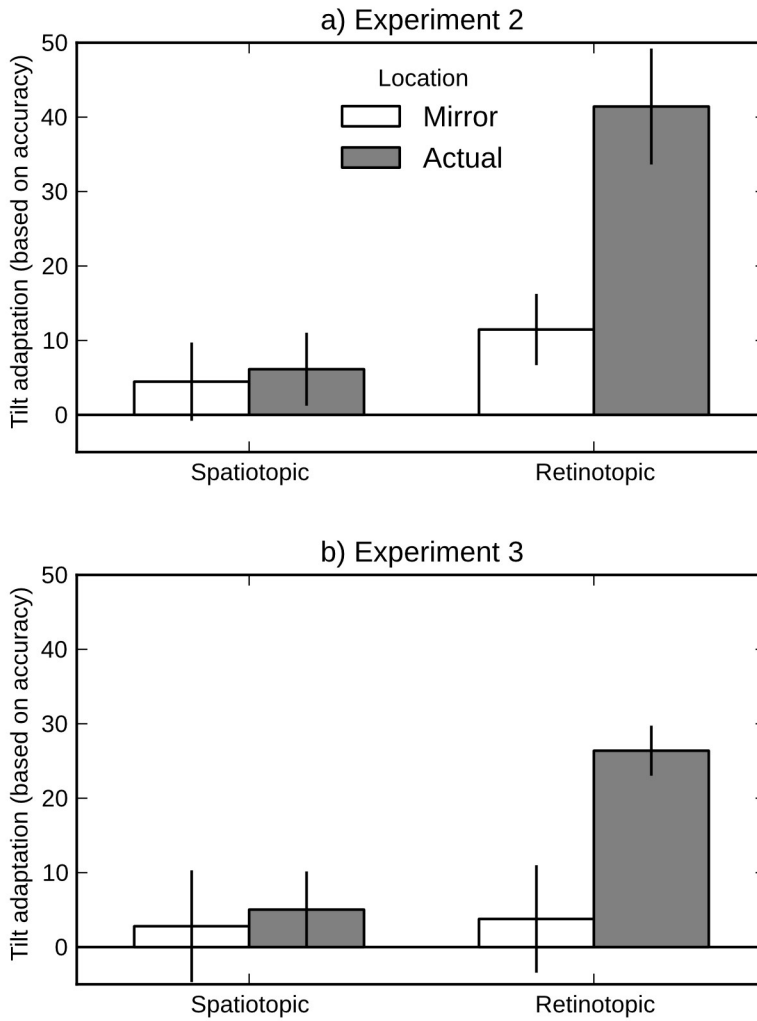
Figure vi.3. Results of Experiments 2 (a) and 3 (b). Although there is some generalised, non-location-specific tilt-adaptation aftereffect (TAE) throughout the visual field, the location-specific TAE is confined to the retinotopic location.

not to locations in space (spatiotopic). However, one might argue that it takes some time for spatiotopic TAE to emerge after a saccade, and that the post-saccadic delay of 100ms was too brief (cf. Mathôt & Theeuwes, 2010b). We therefore repeated the experiment, but presented the tester at a longer interval after the eye movement.

Experiment 3

The aim of Experiment 3 was twofold. Firstly, we wanted to replicate our finding that location-specific TAE is retinotopic. Secondly, we wanted to exclude the possibility that the lack of any spatiotopic TAE in Experiment 2 was due to the fact that the tester was presented fairly rapidly after the eye movement. Therefore, in this experiment, we increased the delay between the onset of the saccade and the presentation of the tester to 500ms, so that the tester was presented far outside the 'window of stability'.

Method

The method was identical to that of Experiment 2, with the following exceptions. 5 new observers participated in the experiment. All were naïve as to the purpose of the study and none were trained psychophysical observers. The tester stimulus was presented 500ms after the onset of a saccadic eye movement had been detected.

Results

Trials were excluded based on the following criteria: The eyes deviated more than 2° from the expected location during the trial (8.4%); Saccade latencies were below 50ms (6.1%) or above 500ms (1.6%); Response times were below 50ms (-) or above 1500ms (1.5%). In total, 82.3% of all trials were included in the analysis.

The same repeated measures ANOVA as in Experiment 2 with TAE_{acc} as dependent revealed an effect of location, $F(1,4) = 20.1$, $p < .05$, and trends toward an effect of condition, $F(1,4) = 5.5$, $p < .1$, and a condition by location interaction, $F(1,4) = 7.0$, $p < .1$ (see Figure vi.3b). These effects were qualitatively identical to those found in Experiment 2. A similar analysis with TAE_{rt} as dependent variable revealed no effects, but yielded qualitatively similar results.

We determined the Bayes factor for the difference between actual and mirror trials in both the spatiotopic and the retinotopic, using the same parameters as in Experiment 2. For the spatiotopic condition, this gave us the following: $M = 2.24$, $SE = 5.31$, $Bf = 0.19$. Again, this indicates "substantial evidence" *against* location-specific TAE. For the retinotopic condition, this gave us the following: $M = 22.61$, $SE = 4.01$, $Bf > 1 \times 10^6$. Again, this indicates "decisive evidence" *in favour of* a location-specific TAE.

Discussion

We replicated the finding that the location-specific component of TAE is retinotopic. In addition, we excluded the alternative explanation that the lack of spatiotopic TAE in Experiment 2 was due to the short interval between saccade onset and tester presentation.

General discussion

The present results strongly suggest that the location-specific component of the tilt-adaptation aftereffect (TAE) is tied to a purely retinotopic frame of reference. Neither shortly (Exp. 2) nor at longer intervals (Exp. 3) following a saccadic eye movement did we find TAE at the originally adapted location. Crucially, by conducting a Bayesian analysis we have shown that this is not a null result due to a lack of statistical power. Our results confirm a recent report of purely retinotopic TAE (Knapen et al., 2010), and are inconsistent with studies that have shown spatiotopic TAE across eye movements (Melcher, 2005, 2008a, 2008b). More generally, our results support the view that there is no 'remapping' of low-level visual features across saccades (Bays & Husain, 2007; Cavanagh et al., 2010; O'Regan & Noë, 2001).

Even in light of the present results, there is no obvious explanation for the fact that some studies have shown spatiotopic adaptation aftereffects across saccades (Ezzati et al., 2008; Melcher, 2005, 2008a, 2008b), whereas other studies, including the present one, have failed to find any such evidence (Afraz & Cavanagh, 2009; Knapen et al., 2009, 2010; Wenderoth & Wiese, 2008). One possibility is that spatiotopic aftereffects may sometimes emerge, but only when attention is focused on the adapter stimulus. This is indirectly supported by a recent functional magnetic resonance imaging study (fMRI) study, in which spatiotopic selectivity was reported in a range of visual brain areas (Crespi et al., 2011). Crucially, this spatiotopic selectivity was found only when attention was focused on the stimulus that elicited the activation, but not under conditions of passive viewing (see also d'Avossa et al., 2007; Gardner et al., 2008). It might be that for some reason, such as subtle differences in the paradigm or instruction, participants sometimes pay attention to the adapter stimulus, while they ignore the adapter in other situations.

An alternative possibility, favoured by Knapen and colleagues (2009), is that generalised, non-location-specific adaptation aftereffects have occasionally been mistaken for a location-specific, spatiotopic effect. This is supported by the observation that, in hindsight, none of the studies that have reported spatiotopic adaptation aftereffects have used a carefully controlled design. In particular, in these studies the spatiotopic and retinotopic selectivity has not been determined by comparing the adaptation effect to

separate control locations, which have been matched in terms of eccentricity relative to both the first and second fixation (e.g., Ezzati et al., 2008), and the direction of the saccade (e.g., Melcher, 2005).

While acknowledging that the issue is open to debate, we believe that the hypothesis that generalised adaptation aftereffects have been mistaken for location-specific, spatiotopic effects is the most parsimonious way to reconcile the divergent findings (cf. Knapen et al., 2010). This also reduces the apparent gap between findings on adaptation aftereffects across saccades and the broader literature on visual stability and trans-saccadic integration. More specifically, the consensus is that detailed, low-level information is mostly, if not entirely, lost across saccades (McConkie & Currie, 1996; McConkie & Zola, 1979), whereas conceptual information is retained to some extent (Henderson et al., 1987), even though the spatial specificity of this form of trans-saccadic integration is debatable (Pollatsek et al., 1990; for reviews, see Irwin, 1996; Mathôt & Theeuwes, 2011a). This also fits with the view, which we have highlighted in Chapter v of this thesis, that spatially specific trans-saccadic integration, or 'remapping', occurs only in the context of action and attention (cf. Bays & Husain, 2007; Cavanagh et al., 2010).

In summary, we have shown that the location-specific component of the tilt-adaptation aftereffect (TAE) is purely retinotopic (see also Knapen et al., 2010). After an eye movement, TAE is not found at the original adapted location, but only at the location that retinotopically matches the adapted location. We have acknowledged that the issue is open to debate, but have suggested that previous reports of spatiotopic adaptation aftereffects have been due to an incorrect choice of control locations, which allowed generalised effects to be mistaken for location-specific, spatiotopic effects (Ezzati et al., 2008; Melcher, 2005, 2008a, 2008b). Finally, we have argued that the present results are consistent with the notion that there is little, if any, trans-saccadic integration of low-level visual information (cf. Bays & Husain, 2007; Cavanagh et al., 2010; Mathôt & Theeuwes, 2011a; O'Regan & Noë, 2001).

∞

VII. IT'S ALL ABOUT THE TRANSIENT: INTRA-SACCADIC ABRUPT ONSETS DO NOT CAPTURE ATTENTION

Abstract — An abrupt onset stimulus was presented while the participants' eyes were in motion. Because of saccadic suppression, participants did not perceive the visual transient that normally accompanies the sudden appearance of a stimulus. In contrast to the typical finding that the presentation of an abrupt onset captures attention and interferes with the participants' responses, we found that an intra-saccadic abrupt onset does not capture attention: It has no effect beyond that of increasing the set-size of the search array by one item. This finding favours the local transient account of attentional capture over the novel object hypothesis.

∞

Contrary to subjective experience, we are driven to a considerable extent by our environment. A prime example of this is attentional capture. When an object suddenly appears, we often cannot help but look at it (Theeuwes et al., 1998). And even if we do manage to avoid directing our gaze towards a novel object, it invariably attracts our attention (Posner, 1980). This involuntary shift of attention can be measured easily in an experimental setting: If a suddenly appearing stimulus (an 'abrupt onset') is not relevant to the task at hand, its appearance will disrupt performance (Theeuwes, 1994). Conversely, if an abrupt onset happens to be task-relevant, even if just by chance, performance will improve (Yantis & Jonides, 1984).

In the present study, we investigate the effect of an abrupt onset that is presented during a saccadic eye movement. Visual perception is greatly impaired during saccadic eye movements, a phenomenon that is generally referred to as saccadic suppression (Matin, 1974). The origin of saccadic suppression is a topic of debate, centred around two main hypotheses (for a recent review, see Castet, 2010). The traditional, and widely accepted, view holds that contrast sensitivity is reduced during eye movements, particularly for low spatial frequencies (Burr & Ross, 1982; Volkmann, Riggs, White, & Moore, 1978).

However, an alternative view holds that saccadic suppression is, at least in part, due to temporal masking (Campbell & Wurtz, 1978; Castet, 2010). According to the masking account, we are not aware of the intra-saccadic percept, because it is masked by the post-saccadic image. However, although the origin of saccadic suppression is topic of debate, its perceptual effect is clear: Visual events that occur while the eyes are in motion are not perceived (except under rare circumstances, cf. Castet & Masson, 2000). For the purpose of the present experiment, saccadic suppression is therefore a useful tool, because it allows us to present an abrupt onset, while preventing participants from perceiving the exact moment of its appearance (the 'visual transient').

Whether or not an intra-saccadic abrupt onset captures attention speaks to the mechanism that underlies attentional capture. Some authors have suggested that novel objects capture attention by virtue of being novel perceptual entities (analogous to object-files, cf. Kahneman, Treisman, & Gibbs, 1992), regardless of whether their appearance is accompanied by a visual transient (the novel object hypothesis). Support for this view comes mostly from experiments in which stimuli were equiluminant with the display background (Yantis & Hillstrom, 1994; Yantis & Jonides, 1996; also see Davoli, Suszko, & Abrams, 2007). These experiments have shown that a luminance increment is not required for objects to capture attention, in apparent support of the novel object hypothesis. With regard to the present study, the novel object hypothesis predicts that intra-saccadic abrupt onsets capture attention, because they are novel perceptual objects even though the visual transient of their appearance is not perceived.

Others have argued that attentional capture is solely driven by visual transients (the local transient hypothesis; Franconeri, Hollingworth, & Simons, 2005; Hollingworth, Simons, & Franconeri, 2010; Jonides & Yantis, 1988). Evidence that favours the local transient hypothesis over the novel object hypothesis comes predominantly from paradigms in which a novel object appears without being accompanied by a unique visual transient. For example, in a cleverly designed experiment, Franconeri and colleagues (2005) used a contracting annulus that briefly occluded a search array by moving over it. When the search array re-emerged from underneath the annulus, it included a novel object. Their crucial finding was that the novel object did not capture attention when its appearance was hidden from view by the annulus (but see Chua, 2009).

There appears to be an implicit, but important assumption that underlies much of the debate between the novel object and local transient accounts of attentional capture: the assumption that a luminance increment is the only transient that matters (e.g., Jonides & Yantis, 1988; Yantis & Hillstrom, 1994; Yantis & Jonides, 1996). A different, more

contemporary view, which underlies most models of visual attention (Itti & Koch, 2001; Itti, Koch, & Niebur, 1998), is that distinct features (orientation, colour, etc.) are processed in parallel (Koch & Ullman, 1985). In this view, a change from one orientation to another, or from one colour to another, constitutes a visual transient, just like a luminance increment. Therefore, in order to adequately test the novel object account of attentional capture, one would need to present a novel object without a visual transient of any kind, not just without a luminance increment. In our view, the only way to achieve this is by presenting a novel object during an eye movement, while visual perception is greatly impaired (Matin, 1974).

At present, evidence favours the view that intra-saccadically presented stimuli do not capture attention. This, in turn, supports the local transient account of attentional capture. Crucial in this regard are the classic studies on changes blindness (Rensink, O'Regan, & Clark, 1997; Grimes, 1996; for similar findings, see Mack & Rock, 1998; Bridgeman, Hendry, & Stark, 1975). In one variation of the change blindness paradigm, two images are presented in alternation (Brockmole & Henderson, 2005a, 2005b; Grimes, 1996). Both images are identical, except for the presence of a particular object. For example, a tree may be present in one image, but not the other. In the experimental condition, the moment at which the additional object appears is time-locked to the onset of a saccadic eye movement. In the control condition, the object appears during fixation. The typical finding is that people are very poor at detecting the novel object when it appears during an eye movement, whereas detection is exceedingly efficient when the object appears during fixation. The explanation is that normally a suddenly appearing object constitutes a visual transient that captures our attention. But when the change occurs during saccadic suppression, the visual transient is not perceived, and the change no longer captures our attention.

However, it is important to clearly define the limits of what we can conclude from these findings. Change blindness experiments elegantly show that intra-saccadic abrupt onsets have, at most, a small effect when studied in the context of natural scenes. But they do not show that intra-saccadic abrupt onsets have no effect at all, even when using a paradigm that has been specifically designed to elicit a robust effect of attentional capture.

More specifically, the scenes that were used in previous studies contained many objects, which were not organised in any obvious pattern (Brockmole & Henderson, 2005a, 2005b). This made it difficult for participants to infer whether a new object had appeared. In contrast, in the current experiment we initially presented only three objects, with the novel intra-saccadically presented object being the fourth. Furthermore, the

display was highly structured. The small set size and high degree of structure made it possible for participants to retain the entire search display in trans-saccadic/ visual working memory, which has an estimated capacity of about four objects (Irwin, 1992b; Luck & Vogel, 1997; Prime et al., 2007). Consequently, compared to previous studies, it was very easy for participants to infer the appearance of the abrupt onset. The crucial question is whether attentional capture is reinstated under these circumstances.

Another possibility is that intra-saccadic abrupt onsets capture attention only when they occur in a retinotopic frame of reference. This may seem like a quixotic hypothesis, but recent studies have shown that, immediately after an eye movement, attentional effects are, in some cases, predominantly retinotopic (Golomb, Chun, & Mazer, 2008; Golomb, Marino, Chun, & Mazer, 2011; Mathôt & Theeuwes, 2010a, 2010b; but see Pertzov, Zohary, & Avidan, 2010). Presumably, this is because every eye movement is followed by a brief 'window of instability' during which the visual system has not yet fully updated its retinotopic representation (Morris et al., 2012; for a review see Mathôt & Theeuwes, 2011a).

Therefore, with the present study we aim to answer the following questions: First, do intra-saccadic abrupt onsets completely fail to capture attention, even in a paradigm that has been designed to elicit a robust effect of attentional capture? And second, in which reference frame do intra-saccadic abrupt onsets capture attention, if they capture attention at all?

Experiment 1

The aim of Experiment 1 was to investigate whether a suddenly appearing stimulus captures attention when the onset of the stimulus occurs during an eye movement (an intra-saccadic abrupt onset). We used a variation of the attentional capture paradigm, introduced by Theeuwes (1994; for a recent review, see Theeuwes, 2010).

In a typical attentional capture paradigm, participants report the orientation of a line-segment in a unique placeholder (e.g., a single red circle among multiple green circles). Initially, all potential target line-segments are masked. On a proportion of trials, an additional stimulus (an abrupt onset) is presented simultaneously with the unmasking of the target (i.e., the moment when the target becomes visible). Empirically, the typical result is that participants are slower to respond when an abrupt onset is presented. Theoretically, the interpretation is that the abrupt onset captures the participants' attention, diverting attention away from the target stimulus, thus delaying their response.

In our variation of this task, the unmasking of the search array and the presentation of the onset were time-locked to the onset of a saccadic eye movement. Furthermore, there

were three conditions, in order to determine the reference frame of the attentional capture effect, if any.

In the *spatiotopic* condition, the masked and unmasked search arrays were presented at the same location on the display. This allowed us to investigate whether an intra-saccadic abrupt onset captures attention when it occurs in spatiotopic, or world-centred, coordinates.

In the *retinotopic* condition, the masked and unmasked search arrays were presented in the same retinal coordinates: The search array moved with the eyes. This allowed us to investigate the effect of an abrupt onset in retinotopic, or eye-centred, coordinates.

Finally, in the *both* condition, the abrupt onset was both a retinotopic and a spatiotopic visual event. This allowed us to investigate whether the effects of intra-saccadic abrupt onsets (if any) in the retinopic and spatiotopic conditions would be additive, or interact in some way.

To avoid the possibility of overloading trans-saccadic/ visual working memory, which has an estimated capacity of about 4 items (Irwin, 1992b; Luck & Vogel, 1997; Prime et al., 2007), we used a set-size of 3 (not including the abrupt onset, cf. Theeuwes, 1994).
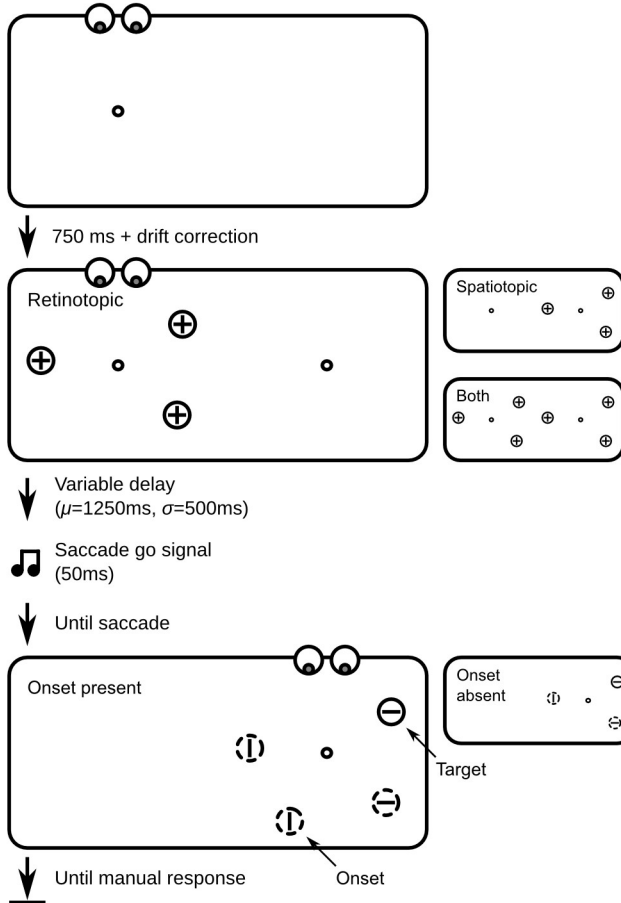
Methods

8 observers, including one of the authors (SM), participated in the experiment. All participants were between the ages of 21 and 39, and reported normal or corrected visual acuity. Eye movements were recorded using an Eyelink 1000 (SR Research, Mississauga, ON, Canada), a video based eye tracker sampling at 1000Hz. The experiment was created using OpenSesame (Mathôt, Schreij, et al., 2012). Stimuli were presented on a 22" CRT monitor, with a resolution of 1024x768px and a refresh rate of 100Hz.

A schematic example trial is shown in Figure vii.1a. Before the start of each trial, a white fixation dot (r=0.25°) was presented against a dark background. This fixation dot was presented 9.4° to the left or the right of the display centre. After 750ms, which gave participants sufficient time to (re)fixate, a drift correction procedure was executed. Drift correction was triggered when a sustained fixation was detected and did not require a manual response from the participant.

Next, the trial proper started with the presentation of a masked search array and a saccade target that was identical to the fixation dot. The saccade target was always presented at the mirror location of the fixation dot. The masked search array consisted of three unfilled circular placeholders (r=1.06°). All placeholders were of the same colour, which could be either red or green. All placeholders contained a cross, consisting of a white vertical and horizontal line-segment (1.25°). All placeholders were presented at

Figure vii.1. a) A schematic example trial of Experiment 1 in the retinotopic, onset present condition. The target is defined as the line-segment in the uniquely coloured circle (different colours are indicated by dashed versus solid lines). The abrupt onset appears at a previously unoccupied location, either in retinotopic coordinates (eye-centred; as shown here), spatiotopic coordinates (world-centred), or both. b) A schematic example trial of Experiment 2 in the set size 3, onset absent condition.

randomly selected locations, 5.3° from the fixation dot and/ or saccade target (see below), and spaced apart by multiples of 60° angular.

In the *spatiotopic* condition, the masked search array was presented around the saccade target. In the *retinotopic* condition, the masked search array was presented around the fixation dot. In the *both* condition, two identical masked search arrays were presented simultaneously around both the fixation dot and the saccade target (i.e. six placeholders were presented in total).

The participants were instructed to keep fixating on the fixation dot until after a random interval (μ=1250ms, σ=500ms; minimum=500ms) an auditory go-signal was presented (50ms, 440Hz, sine wave). Participants were instructed to make a saccade to the saccade target when they heard the go-signal, but to avoid anticipatory responses.

As soon as a saccade was detected, defined as the moment at which the horizontal gaze position deviated more than 1.9° from the fixation dot (μ=906ms, σ=219ms; unfiltered across all participants), the (unmasked) search array was presented. All placeholders, except the one that contained the target line-segment, changed colour (i.e., from red to green or vice versa). In all placeholders one of the line-segments disappeared, so that a single horizontal or vertical line-segment remained. There was always a single, unmasked search array, which was presented around the saccade target. The *retinotopic*, *spatiotopic*, and *both* conditions differed only in the location and number of the pre-saccadic, masked search array(s).

In the *onset present* condition, an additional distractor, identical to the other distractors, appeared at a randomly selected location that was previously unoccupied. In the *onset absent* condition, no additional distractor was presented.

Participants reported the orientation of the target line-segment (i.e., the line-segment in the uniquely coloured placeholder) as quickly as possible. If the target line-segment was horizontal they pressed the left button on a serial response box (Psychology Software Tools, Sharpsburg, PA, USA). The right button was pressed on a vertical target line-segment.

Onset presence (*absent*, *present*), condition (*spatiotopic*, *retinotopic*, *both*), and initial fixation condition (*left*, *right*) were mixed within blocks. Target colour (*red*, *green*) was fully randomized. The experiment consisted of 36 practice trials, followed by 288 experimental trials.

Results

Trials in which gaze deviated more than 3° from the expected point of gaze (10.3%), trials in which saccade latency was below 50ms (-) or above 1000ms (19.4%)[13], and trials in which response time was below 100ms (0.1%) or above 2000ms (0.4%) were discarded. The mean error rate across valid trials was 10%. Across valid and correct trials, the mean response time was 743ms, and the mean saccade latency was 840ms. An alpha level of .05 was used for all statistical tests. *F*-values are reported when larger than 1.

A repeated measures analysis of variance (ANOVA) was conducted with onset presence (*absent*, *present*) and condition (*spatiotopic*, *retinotopic*, *both*) as within-subject factors and mean correct response time as dependent variable (Figure vii.2a). This revealed an effect of onset presence, $F(1,7) = 7.1$, $p < .05$, such that response times were faster when an onset was presented than when no onset was presented.

A similar repeated measures ANOVA using error rate as dependent variable revealed no significant effects. Tentatively, there was a trend towards a reduced error rate in the presence of a distractor, $F(1,7) = 3.2$, analogous to the effect that was found in the response times.

Discussion

The results of Experiment 1 were unanticipated: The presentation of an abrupt onset caused facilitation, as measured by decreased response times and, tentatively, decreased error rates. This effect, which was present across all three conditions, contrasts with the typical finding that an abrupt onset causes interference. A possible explanation for this paradoxical facilitation was investigated in Experiment 2.

Experiment 2

The aim of Experiment 2 was twofold. Firstly, we wanted to verify our methodology by replicating the typical finding that an abrupt onset causes interference. To do this, we created an experiment that was similar to Experiment 1, but did not involve an eye movement. Secondly, we wanted to investigate the possibility that the paradoxical facilitation observed in Experiment 1 could be explained by a set-size effect. If the visual system disregards all (or most) changes that occur during a saccade, we might expect the visual system to process the post-saccadic search array simply as it is found 'on arrival'. If

---

13 Our instruction emphasised saccadic accuracy over speed. This resulted in the exclusion of a substantial proportion of trials (19.4%) in which participants executed exceedingly slow eye movements (>1000ms). However, there is no reason to suspect that this has systematically influenced the pattern of results highlighted here.

Figure vii.2. a) Results of Experiment 1. Participants were faster to respond if an intra-saccadic abrupt onset was presented. This 'paradoxical facilitation' was found in all three conditions (retinotopic, spatiotopic, and both). b) Results of Experiment 2. Participants were slower to respond if an abrupt onset was presented during fixation, reflecting the typical pattern of attentional capture. Importantly, if no abrupt onset was presented, participants were faster in the set size 4 condition than in the set size 3 condition. c) The set size effect of Experiment 2 fully accounted for the paradoxical facilitation of Experiment 1.

this is the case, in Experiment 1 presenting an intra-saccadic abrupt onset was equivalent to increasing the set-size from 3 to 4.

At first glance you might expect that an effect of set-size, if any, should be such that an increased set-size leads to increased response times (cf. the classic study by Treisman & Gelade, 1980). But previous research has shown that this is not always the case. Under particular circumstances, increasing the number of non-targets will cause the target element to become more salient, resulting in faster response times (Green, 1991, 1992; Sagi & Julesz, 1987). This is particularly the case when the target is defined as the unique stimulus, which is the case in the present experiments. Simply put, a red circle is more unique, and therefore more conspicuous, among three green circles than among two green circles.

To test whether a set-size effect could account for the paradoxical facilitation found in Experiment 1, we included set-sizes of 3 and 4 (not including the abrupt onset) in Experiment 2.

Method

The method was identical to that of Experiment 1, with the following exceptions. 8 observers, including one of the authors (SM), participated in the experiment. All participants were between the ages of 18 and 45, and reported normal or corrected visual acuity. Eye movements were not monitored. Stimuli were presented on a 19" CRT monitor, with a resolution of 1024x768px and a refresh rate of 120Hz.

A schematic example trial is shown in Figure vii.1b. The fixation dot was presented at the centre of the display, and both the masked and unmasked search array were presented centrally, around the fixation dot. In order to match the paradigm as closely as possible to that of Experiment 1, a dummy sound, identical to the saccade go-signal of Experiment 1, was presented. In addition, the saccade latencies of Experiment 1 were 'played back': On each trial, a saccade latency was randomly selected from the valid trials of Experiment 1 and used for the delay interval after the dummy sound.

A set-size of 4 was included in addition to the original set-size of 3 (set-sizes do not include the abrupt onset). Participants reported a horizontal target line-segment by pressing the 'z'-key on a standard keyboard, and a vertical line-segment by pressing the slash-key. Onset presence (*absent*, *present*) and set-size (*3*, *4*) were mixed within blocks. The experiment consisted of 12 practice trials, followed by 192 experimental trials.

Results

Trials in which response time was below 100ms (-) or above 2000ms (0.6%) were discarded. The average error rate across valid trials was 10%. Across valid and correct trials, the mean response time was 747ms.

A repeated measures ANOVA was conducted with onset presence (*absent*, *present*) and set-size (*3*, *4*) as within-subject factors and mean correct response time as dependent variable (Figure vii.2b). This revealed an effect of onset presence, $F(1,7) = 6.7$, $p < .05$, reflecting the typical distractor interference effect. Tentatively, there were trends towards an effect of set-size, $F(1,7) = 3.2$, and a set-size by onset presence interaction, $F(1,7) = 3.1$.

A similar repeated measures ANOVA using error rate as dependent variable revealed an effect of onset presence, $F(1,7) = 16.4$, $p < .01$, again reflecting a distractor interference effect. Tentatively, there was a trend towards an effect of set-size, $F(1,7) = 4.4$.

To get a better estimate of the pure set-size effect in the absence of an abrupt onset, we performed a two-tailed paired samples *t*-test between set-sizes 3 and 4 using only the distractor absent trials, with mean correct response time as dependent variable. This revealed that responses were slower for set-size 3 than for set-size 4, $t(7) = 4.4$, $p < .005$. A similar *t*-test using error rate as dependent variable showed no effect.

Discussion

In Experiment 2 we replicated the typical finding that an abrupt onset stimulus leads to increased response times and error rates, thus boosting our confidence in the overall methodology and results of Experiment 1.

Furthermore, the results show that participants respond faster in the set-size 4 condition than in the set-size 3 condition, at least when no onset stimulus is presented. This confirms the existence of a set-size effect (Green, 1991, 1992; Sagi & Julesz, 1987). In the next section, we will examine whether the set-size effect found in Experiment 2 can account for the paradoxical facilitation observed in Experiment 1.

Cross-experimental analysis

Although it is clear that the paradoxical facilitation (Exp. 1) can be explained at least in part by a set-size effect (Exp. 2), it is not obvious that these two effects completely cancel each other out. To investigate this more rigorously, we conducted a cross-experimental analysis.

First, we derived the magnitude of the set-size effect from the onset absent trials of Experiment 2. We subtracted the mean correct response time in the set-size 3 condition from that in the set-size 4 condition, which gave us a estimated effect magnitude of 31ms.

We did the same for the mean error rate (even though there was no significant set-size effect in the error data) and obtained an effect magnitude of 1.2%.

Next, we added the estimated effect magnitude to the response times and error rates of the onset-present trials of Experiment 1 (Figure vii.2c). Using the 'set-size corrected' data, we performed the same repeated measures ANOVA as before, with onset presence (*absent*, *present*) and condition (*spatiotopic, retinotopic, both*) as within-subject factors and mean correct response time as dependent variable. This revealed no effects, nor did a similar analysis using mean error rate as dependent variable. However, to avoid drawing conclusions from a null-result, we conducted a complementary Bayesian analysis.

We determined the size of the attentional capture effect in the set-size corrected data, by taking the difference in mean response time between onset present and onset absent trials, collapsed over the three conditions. This gave us the following estimate: $M = -1.6$ms, $SE = 13.3$ms. Next, we chose realistic lower and upper bounds for the effect. We set the lower bound to 0ms, since very small yet reliable capture effects have been reported (e.g., ~10ms reported by Mulckhuyse, Talsma, & Theeuwes, 2007). We set the upper bound to 100ms, since very large capture effects tend to be in this range (e.g., ~85ms for small separation trials in Exp. 2 of Mathôt & Theeuwes, 2011b). Using these parameters, and assuming a uniform prior distribution, we determined the Bayes factor (cf. Dienes, 2011), $Bf = 0.15$. Following Jeffreys (1961; reproduced in Wetzels et al., 2011), this constitutes substantial evidence for H_0.

In summary, there is substantial evidence that the paradoxical facilitation (Exp. 1) is fully accounted for by a set-size effect (Exp. 2).

General discussion

In the present study we pitted the novel object account of attentional capture (Chua, 2009; Davoli et al., 2007; Yantis & Hillstrom, 1994; Yantis & Jonides, 1996) against the local transient account (Franconeri et al., 2005; Hollingworth et al., 2010; Jonides & Yantis, 1988). We presented a stimulus during a saccadic eye movement. Because visual perception is greatly impaired while the eyes are in motion (Matin, 1974), this manipulation masked the visual transient that normally accompanies the appearance of a stimulus. Phrased differently, in our experiments participants did not perceive the exact moment at which the stimulus appeared.

The main finding is that an intra-saccadically presented stimulus does not capture attention—it has no effect beyond that of increasing the set size of the search display by one. By most definitions, a suddenly appearing stimulus is a novel perceptual entity, regardless of whether its appearance is masked by saccadic suppression. Therefore, this

result strongly favours the local transient account, which postulates that a visual transient is required for attentional capture to occur, over the novel object account.

The present study complements previous research in a number of important ways. Franconeri and colleagues (2005) found that a novel stimulus does not capture attention when it emerges from underneath an occluder. Based on this finding they concluded that novel objects do not capture attention without a visual transient, in line with our own views. However, the presence of a moving occluder might have had unanticipated side-effects, for example because it is likely to capture attention itself (Franconeri & Simons, 2003b). In contrast, our manipulation, which was one of timing, did not interfere with the search display in a similar way. Our results therefore offer important corroborative evidence for the crucial role of visual transients in attentional capture.

Furthermore, we used a sparse and structured display, which consisted of three to four items. This is a critical departure from previous studies on intra-saccadically presented stimuli, which have generally used unstructured and complex natural scenes (e.g., Brockmole & Henderson, 2005a, 2005b). It has been shown that the capacity of visual working memory is about 4 items (Luck & Vogel, 1997). Given the involuntary allocation of attention to the target of an upcoming eye movement (e.g., Deubel & Schneider, 1996; Godijn & Pratt, 2002), and the tight coupling between attention and working memory (Awh & Jonides, 2001), it is conceivable that the effective capacity of working memory is somewhat reduced in paradigms that require an eye movement, such as the one used here. Yet, previous studies have consistently shown that such impairment, if any, is negligible, and that trans-saccadic memory (i.e. working memory across saccades) has a comparable capacity of 3 to 4 items (Irwin, 1992b; Prime et al., 2007; for a recent review, see Prime, Vesia, & Crawford, 2011). Therefore, in our experiments the entire search array could be retained in memory. The finding that, even under these circumstances, an intra-saccadically presented novel object does not capture attention, strongly suggests that a visual transient is indeed required for attentional capture to occur: Previous failures to find attentional capture by intra-saccadically presented stimuli cannot be attributed to an overload of working memory.

In defence of the novel object account of attentional capture, one could argue that the perceptual effect of a saccade is not unlike that of a sudden visual disruption (Castet, 2010). In that sense, our strategy of presenting a sudden onset during a saccade is comparable to previous experiments, in which a sudden onset was presented during a visual disruption (Davoli et al., 2007; Franconeri et al., 2005; Hollingworth et al., 2010). Consequently, one could argue that the present experiment was not a fair test of the novel

object hypothesis, because object representations might not be robust to the disruptions caused by saccadic eye movements.

However, the crucial distinction between the local transient and novel object hypothesis is the level at which attentional capture is assumed to operate. The local transient account assumes that attentional capture is triggered by a sudden change in some low-level feature of the visual input, such as a luminance increment or colour change. In contrast, the novel object hypothesis postulates the existence of perceptual entities (analogous to object-files, cf. Kahneman et al., 1992), which are, in some sense, detached from low-level visual input. The term 'perceptual entity' is not clearly defined, but it is presumed to reflect some "relatively high-order mechanism" (Yantis & Hillstrom, 1994, p. 106) that "bridges over the discontinuities produced by (...) saccades" (Kahneman et al., 1992, p. 178). Given these definitions, the results of the present study constitute clear evidence against the novel object hypothesis.

The present results also speak to the mechanisms that underlie visual stability—our ability to integrate visual information from one fixation to the next. In a classic essay, MacKay (1972) suggested that eye movements can be viewed as questions. Before every eye movement we 'ask' whether, after the eye movement, all stimuli will be where we expect them to be. The answer is provided by the image that is seen after the eye movement. If the answer does not match the question, a change is detected (this general idea, often referred to as 'state feedback control', is also increasingly applied outside of the domain of vision, see e.g. Hickok, 2012). Crucially, if a stimulus is presented only after the eyes have set in motion, as we have done here, it will not be part of the 'question'. Consequently, the change is not automatically detected, and we have to resort to higher level processes (Brockmole & Henderson, 2005a). Using more contemporary terminology, visual stability relies on a sparse representation that includes only the objects that were in the focus of attention prior to the eye movement (Irwin, 1991, 1996; Prime et al., 2007; for a review, see Mathôt & Theeuwes, 2011b). Because an intra-saccadic abrupt onset appears at a previously unattended location and does not capture attention by itself, its sudden appearance has little effect on our subjective sense of visual stability and on our behaviour. This is demonstrated particularly saliently by the finding that presenting an intra-saccadic abrupt onset is equivalent to increasing the set size by one item: Apparently, after the eye movement the image is evaluated more or less anew, without regard to whether the abrupt onset was present before the eye movement or not.

Finally, we have shown that the lack of attentional capture by intra-saccadically presented objects cannot be explained by the fact that just after an eye movement the

visual system has not yet fully updated its retinotopic representation (Golomb et al., 2008; Mathôt & Theeuwes, 2010a, 2010b; Morris et al., 2012): Retinotopically matched abrupt onsets do not capture attention any more than spatiotopically matched abrupt onsets, or a combination of both.

In summary, with a paradigm designed to elicit a robust effect of attentional capture, the present study confirms that intra-saccadically presented stimuli do not capture attention (Brockmole & Henderson, 2005a, 2005b; Grimes, 1996). It is all about the transient.

<div align="center">∞</div>

VIII. FROM REORIENTING OF ATTENTION TO BIASED COMPETITION

Abstract — When a distractor was presented simultaneously with, or directly following a target, it produced more interference when it was presented in the same visual hemifield as the target, relative to when it was presented in the opposite visual field. This result is interpreted in terms of biased competition: There is more competition between stimuli when they are presented in the same visual field, relative to opposite visual fields. However, when the distractor was presented 125ms or more prior to the target this pattern reversed. In those cases there was more distractor interference when target and distractor were presented in opposite visual fields. This can be explained by assuming that attention was captured by the distractor, and that there was an additional cost of reorienting to a location in the opposite visual field.

∞

Processing of visual information is such that, as impulses travel upwards in the hierarchy of visual brain areas, increasingly complex visual features are processed in a decreasingly localised manner (Smith, Singh, Williams, & Greenlee, 2001). Because of the coarse localisation in higher visual areas, stimuli that are presented near one another activate largely overlapping neural populations. It is therefore difficult to attribute neural activity unambiguously to a single stimulus (Desimone & Duncan, 1995; Luck, Girelli, McDermott, & Ford, 1997; Moran & Desimone, 1985): The spiking of a neuron might indicate that one of the stimuli in its receptive field (RF) has a certain feature, but which stimulus?

It has become clear that visual attention plays a crucial role in the way these ambiguities are resolved. In a situation where attention is unfocused, two nearby stimuli act in a mutually suppressive way (Kastner, De Weerd, Desimone, & Ungerleider, 1998), so that both stimuli are weakly represented. Furthermore, the representations of the two stimuli are not clearly delineated, so that there is ambiguity in which features belong to

which object (for a review, see Beck & Kastner, 2005). However, if attention is directed to one of the stimuli, activation is driven largely by the attended stimulus and the visual system effectively becomes 'blind' to the presence of the non-attended stimulus (Luck, Chelazzi, Hillyard, & Desimone, 1997; Motter, 1993; Reynolds, Chelazzi, & Desimone, 1999). Therefore, in a state of focused attention, the attended stimulus is strongly represented and ambiguity is resolved, since only features of a single object—the attended object—are represented. Phrased differently, nearby stimuli activate overlapping neural populations and therefore compete for representation in the visual system. Visual attention is what determines which stimulus wins this competition. These competitive interactions form the central tenet of many contemporary theories of attention, such as the biased competition model (Desimone, 1998; Desimone & Duncan, 1995), the ambiguity resolution theory (Luck, Girelli, et al., 1997) and the selective tuning model of visual attention (Tsotsos et al., 1995).

As stimuli are spaced further apart there is decreasing overlap between the neural populations that are activated by their presentation. Since competition is believed to occur predominantly at the level of RFs, this results in less competition for representation (Beck & Kastner, 2005; Hopf et al., 2006; but see Tombu & Tsotsos, 2008). In addition, and of special importance to the present study, is the finding that there is also less competition if stimuli are presented in opposite visual fields (Sereno & Kosslyn, 1991; Torralbo & Beck, 2008). In part, this can be explained by the fact that the RFs of neurons in early visual areas tend to be confined to one visual field: The left visual field is processed in the right hemisphere and vice versa (Desimone & Schein, 1987; Gattass, Sousa, & Gross, 1988). However, in the inferior temporal cortex the majority of neurons have RFs that encompass part of both visual fields (Rocha-Miranda, Bender, Gross, & Mishkin, 1975) and even in this area there is more competition if stimuli are presented in the same visual field, compared to if they are presented in opposite visual fields (Chelazzi, Duncan, Miller, & Desimone, 1998; Sato, 1988, 1989).

One of the most important predictions that can be derived from biased competition is that interference between stimuli should be strongest if stimuli are presented in proximity and within the same visual field. That this is the case has been known for quite some time (Pan & Eriksen, 1993; Sereno & Kosslyn, 1991), but not until recently has the link with biased competition been made explicit (Bahcall & Kowler, 1999; Caputo & Guerra, 1998; Cave & Zimmerman, 1997; Mounts, 2000a, 2000b; Torralbo & Beck, 2008). Mounts and Gavett (2004) investigated the effects of biased competition in a behavioural paradigm. Participants were presented with an array of placeholders, followed by an array of stimuli,

and performed a discrimination task using one of these stimuli (the target). Just before the search array was presented, two placeholders suddenly increased in size. The target was always presented at the location of one of these two placeholders. The crucial finding was that participants were faster to respond if the two possible target locations were spaced further apart. Importantly, on top of this distance effect there was an effect of visual field: Participants were faster to respond if the two locations were on opposite sides of the vertical meridian. These findings are readily explained in terms of biased competition: the sudden increase in size of the two placeholders captured attention, thus facilitating processing of the stimuli that were subsequently presented at these locations. However, the extent to which the two stimuli interfered with each other depended on the level of competition between them. Since competition is strongest between stimuli presented in proximity and in the same visual field, there was less interference (as measured by decreased reaction times) if they were spaced far apart and located in opposite visual fields.

Although the implications of biased competition for simultaneously presented stimuli are quite clear, less is known about the temporal characteristics of these competitive interactions. Therefore, an important question is how competition changes as a small delay is introduced between the presentation of two stimuli. A number of studies (Beck & Kastner, 2005; Kastner et al., 2001, 1998) have shown that competitive interactions are absent or strongly reduced if stimuli are presented sequentially. However, the intervals used in these studies are relatively long (250ms between successive presentations) and the possibility remains that competition will occur when a target and a distractor are presented in close temporal succession. It is known that the neural response to a stationary object has a rapid initial build-up and then decreases over time (Keysers & Perrett, 2002; Reynolds et al., 1999). The remaining sustained neural response elicited by a stimulus could compete for representation with another, subsequently presented stimulus. This would result in competition which is less pronounced, but similar to that found in studies where stimuli are presented simultaneously (e.g., Torralbo & Beck, 2008).

The present study followed up on previous research that has shown that competition is strongest between stimuli presented in the same visual field. We investigated how this hemifield effect changes over time, by varying the temporal interval between the presentation of two stimuli. To this end, we used an attentional capture paradigm (cf. Theeuwes, 1994) in which participants responded to a target stimulus in the presence of an irrelevant onset distractor. In an attentional capture paradigm, the typical finding is that the presentation of an irrelevant onset disrupts processing of a target, as measured by increased

reaction times and decreased accuracy relative to a control condition (Theeuwes, 1991). Traditionally, this type of distractor interference has been explained in terms of shifting attention: attention is automatically pulled towards the distractor and the voluntary process of reorienting towards the target takes time and effort (Jonides & Yantis, 1988). However, this view offers no explanation for why distractor interference should be greatest if stimuli are presented in proximity or in the same visual field, at least if they are presented simultaneously. Rather, one might expect the opposite, since shifting attention between two nearby locations intuitively seems less costly than shifting attention across a large region of space. Biased competition offers a related, but more concrete explanation: An irrelevant onset does indeed attract attention, but interference is due to the fact that the target initially "loses the competition" (Mounts, 2000b). Because the strength of competition is inversely related to stimulus separation, there is less competition to be lost, and therefore less distractor interference, if the target and the distractor are spaced far apart or presented in opposite visual fields.

In the present study we presented a target and a distractor either simultaneously or sequentially, and either in the same or in opposite visual fields. As a control condition we included trials in which the distractor was presented not as a single onset, but simultaneously with the premasks (see Figure viii.1a). Our expectation was that when the target and the distractor are presented simultaneously, distractor interference is greater when the distractor and the target are presented in the same visual field, relative to when they are presented at equal distance in opposite visual fields (Mounts & Gavett, 2004; Sereno & Kosslyn, 1991; Torralbo & Beck, 2008). We did not expect competition to arise when there was a significant delay between the presentation of the target and the distractor. However, it is still an open question if competitive interactions also occur, although perhaps to a lesser extent, if the distractor is presented slightly before the target. In Experiment 1 we investigated this by presenting a distractor simultaneously with or prior to the presentation of a target. Another important question is how processing of a target is disrupted by a subsequently presented distractor: do competitive interactions play a role in this? This we investigated in Experiment 2 where a distractor was presented simultaneously with or following the presentation of a target.

Experiment 1

Method

Participants

The observers were 25 students from the Vrije Universiteit Amsterdam, aged 17 through 25. For their participation they received monetary compensation or course credit. All observers reported normal or corrected visual acuity and were naive as to the purpose of the experiment.

Apparatus

The experiment was run on a 2.6GHz Pentium 4 PC, using a 17" non-flat CRT display with a resolution of 1024x768px. Manual responses were collected on a QWERTY keyboard. Three identical computer set-ups were used for experimental testing and colours were adjusted for each display separately. The experiment was conducted in a dimly lit room.

Stimuli

Each trial started with the presentation of a grey central fixation dot (radius 0.2º, luminance 29 cd/ m^2) on a black display (0.5 cd/ m^2). After 700ms, a display consisting of 6 premasks (identically coloured circles containing 6 lines) was presented (Figure viii.1). Circles were presented equidistant from one another (6.4º) on a hypothetical circle with a radius of 12.4º. Each circle had a red (CIE x, y chromaticity coordinates of 0.6, 0.3) or green (0.3, 0.6) outline (0.1º) with a luminance of 12 cd/ m^2. Each circle contained a premask, consisting of 6 grey line-segments (.03º x .9º, 29 cd/ m^2), tilted 0º, 20º, 70º, 90°, 110º, 160º or 180º from a vertical orientation. After 1000ms the search array was presented. All circles except for the target circle changed colour and in each circle all but one line disappeared. In the non-target circles the remaining line was always tilted 20º, 70º, 110º, or 160º from a vertical orientation. In the target circle the remaining line was always oriented vertically or horizontally. In the control condition an additional premask was presented together with the regular premasks (Figure viii.1a). In the other conditions, an additional premask (distractor) was presented at a variable stimulus onset asynchrony (SOA) before, or simultaneously with the presentation of the search array (Figure viii.1b). With the presentation of the search array the additional premask changed in a similar fashion to the other non-target circles.
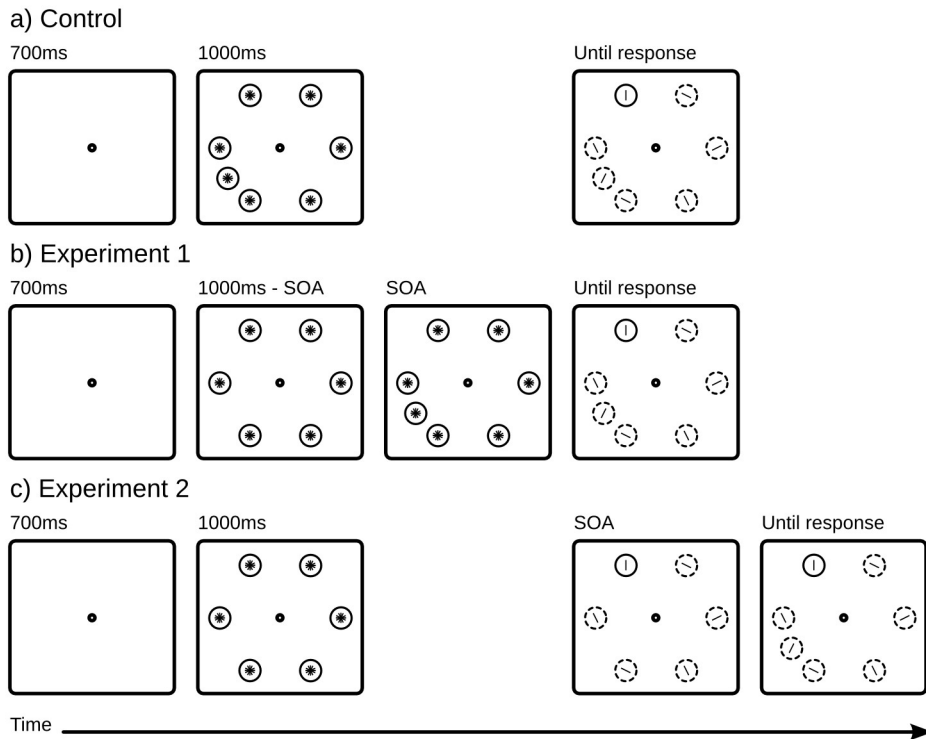
a) Control

b) Experiment 1

c) Experiment 2

Time

Figure viii.1. A schematic representation of the experimental paradigm. Dashed and continuous lines represent different colours. A) The control condition for both experiments, in which the distractor was presented at the same time as the premasks. B) In Experiment 1 the distractor was presented simultaneously with or at a variable SOA before the presentation of the target. C) In Experiment 2 the distractor was presented simultaneously with or at a variable SOA after the presentation of the target.

The distractor was always presented at a distance of 9.5º from the target. In 50% of the trials the distractor was presented in the opposite hemifield from the target, in the remaining trials the distractor and the target were presented in the same hemifield. Neither target or distractor was ever presented on the vertical meridian. SOA (7 levels, including control) and Hemifield (same, opposite) were randomized within blocks.

Procedure and design

Participants received verbal instructions prior to the experiment. Participants were required to make a speeded response of the orientation of the line-segment inside the uniquely coloured circle, by pressing the "Z"-key on presentation of a horizontal line-segment and the "/"-key on a vertical line-segment. It was stressed that gaze should

remain on the fixation point at all times and that accuracy should be around 90%. Only during practice trials did participants receive immediate visual feedback upon an erroneous response. Visual feedback was presented after every block, informing participants of their average RT and the number of errors. The experiment consisted of 48 practice trials and 700 experimental trials. 28 trials constituted a block, after each of which participants were given the opportunity to rest.

Results

One participant was excluded due to overly high RTs (average RT was more than 4 SDs above the grand mean). Trials in which the RT was either below 100ms or above 2000ms (0.04%) were discarded. The error rate across valid trials was 6.2%.

A repeated measures analysis of variance (ANOVA) using Hemifield (same, opposite) and SOA (7 levels, including control) as within-subject factors and mean correct RT as dependent variable revealed a main effect of SOA ($F(6, 138) = 25.59$, $p < .0001$) and an interaction between Hemifield and SOA ($F(6, 138) = 12.30$, $p < .0001$). There was no main effect of Hemifield ($F < 1$). Planned comparisons revealed that in the 0ms SOA condition RTs were slower for distractors presented in the same hemifield as the target ($M = 664$ms, $SE = 20.20$, see Figure viii.2a), compared to distractors presented in the opposite hemifield from the target ($M = 628$ms, $SE = 18.28$; $t(23) = 5.40$, $p < .0001$). The reverse pattern was found in the -125ms SOA condition: RTs were faster for distractors presented in the same hemifield as ($M = 606$ms, $SE = 17.71$), relative to in the opposite hemifield from the target ($M = 627$ms, $SE = 18.81$; $t(23) = 4.56$, $p < .001$). This was also the case for the the -525ms SOA condition ($M = 588$ms, $SE = 17.04$ vs $M = 598$ms, $SE = 17.61$; $t(23) = 2.83$, $p < .01$).

A repeated measures ANOVA using Hemifield and SOA as within-subject factors and the proportion of erroneous responses as dependent variable revealed a main effect of SOA ($F(6, 138) = 5.00$, $p < .001$). This effect was such that accuracy declined as the distractor and the target were presented in closer temporal succession. There were no other effects (all $F < 1$).

Discussion

The results clearly show that when the target and the distractor were presented simultaneously there was less distractor interference when they were presented in opposite visual fields, compared to when they were presented in the same visual field. This effect is consistent with previous studies and is believed to reflect competition for representation between the two simultaneously presented stimuli (Mounts & Gavett, 2004; Sereno & Kosslyn, 1991; Torralbo & Beck, 2008). The influence of competitive interactions declined
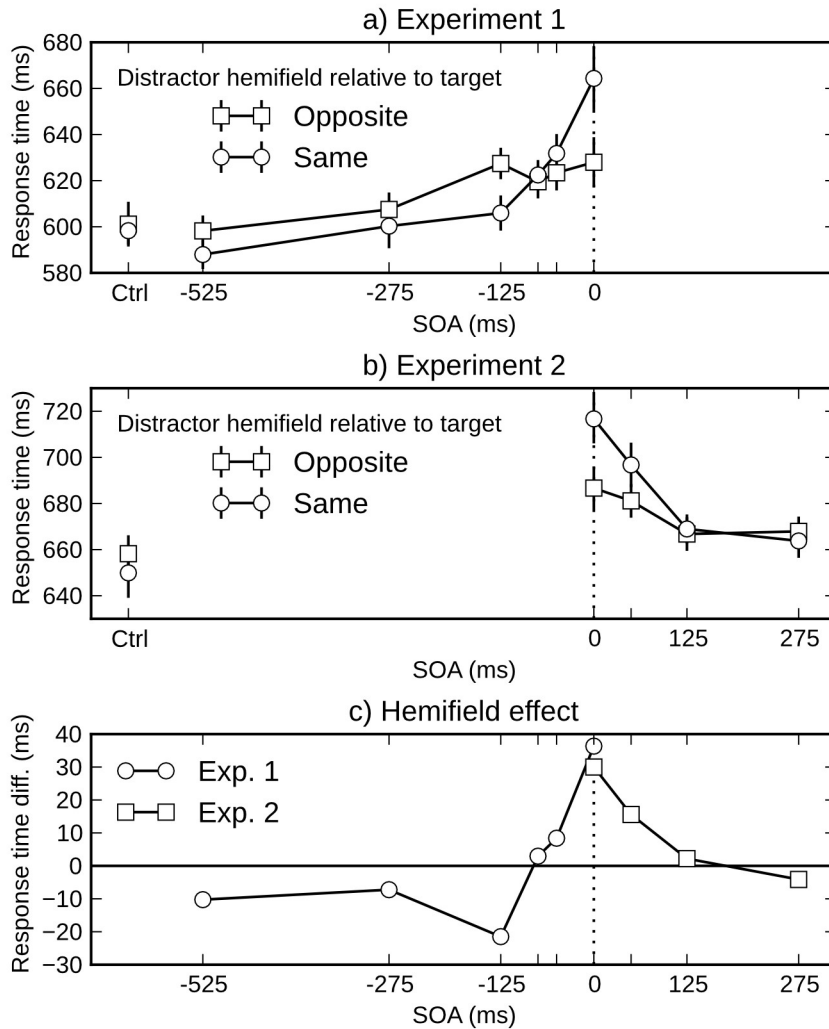
Figure viii.2. a, b) Results of Experiments 1 and 2. When the distractor was presented ≥125ms before the target there was maximal interference when target and the distractor were presented in opposite visual fields. However, when target and distractor were presented (almost) simultaneously there was maximal interference when they were presented within the same visual field. Error bars denote a 95% within-subjects confidence interval as outlined by Cousineau (2005). c) The difference in RT between same visual field and opposite visual field trials, as a function of SOA. Positive values correspond to faster responses on opposite visual field trials, indicative of competitive interactions. Negative values correspond to faster responses on same visual field trials, indicative of preferential reorienting of attention within a visual field.

steeply as a small inter-stimulus interval was introduced and appeared to be complete gone when the distractor was presented 75ms before the target. In addition, we observed a striking pattern reversal when the distractor was presented at longer intervals (125ms or longer) before the target. In those cases, there was more distractor interference when the distractor was presented in the opposite visual field from the target. This can be explained in the more traditional view of reorienting the focus of attention (e.g., Jonides & Yantis, 1988). Because of its sudden onset, attention was captured by the distractor. When at this point the target was presented, attention had to be reoriented towards the target. However, because attention is known to spread preferentially within a visual field (Hughes & Zimba, 1985, 1987; Rizzolatti et al., 1987) it was less costly to reorient attention within a visual field, than it was to reorient attention to a location in the opposite visual field.

Experiment 2

Method

Experiment 2 was identical to Experiment 1 except for the following differences. The observers were 34 students from the Vrije Universiteit Amsterdam, aged 17 through 30. Whereas in Experiment 1 the distractor was presented simultaneously with or prior to the target, in this experiment the distractor was presented simultaneously with or after the presentation of the target, using the following SOAs: 0ms, 50ms, 125ms and 275ms (Figure viii.1c). As in Experiment 1, in the control condition the distractor was presented simultaneously with the premasks. The experiment consisted of 80 practice trials and 480 experimental trials. 40 trials constituted a block.

Results

Trials in which the RT was either below 100ms or above 2000ms (0.08%) were discarded. The error rate across valid trials was 4.2%.

A repeated measures analysis of variance (ANOVA) using Hemifield (same, opposite) and SOA (5 levels, including control) as within-subject factors and mean correct RT as dependent variable revealed a main effect of SOA, $F(4, 132) = 30.75$, $p < .0001$, a main effect of Hemifield ($F(1, 33) = 7.91$, $p < .01$) and an interaction between Hemifield and SOA ($F(4, 132) = 8.92$, $p < .0001$). Planned comparisons revealed that in the 0ms SOA condition RTs were slower for distractors presented in the same hemifield as the target ($M = 717$ms, $SE = 15.16$, see Figure viii.2b), relative to distractors presented in the opposite hemifield ($M = 687$ms, $SE = 14.52$; $t(33) = 5.20$, $p < .0001$). The same pattern pattern was found in the 50ms SOA condition ($M = 697$ms, $SE = 16.15$ vs $M = 681$ms, $SE = 15.47$; $t(33) = 2.82$, $p < .01$).

A repeated measures ANOVA using Hemifield and SOA as within-subject factors and the proportion of erroneous responses as dependent variable revealed a main effect of SOA ($F(4, 132) = 3.38$, $p < .05$), a marginally significant main effect of Hemifield ($F(1, 33) = 3.09$, $p < .1$) and a marginally significant interaction between Hemifield and SOA ($F(4, 132) = 2.32$, $p < .1$). These accuracy effects were such that they mirrored the RT effects.

Discussion

As in Experiment 1, the results provide clear evidence for competition between the target and the distractor when they were presented simultaneously. Importantly, competition was also observed when the distractor was presented 50ms after the presentation of the target. This shows that even if processing of the target has already begun, competitive interactions with a subsequently presented stimulus can be disruptive. At longer SOAs (125ms or more) there was no effect of hemifield. Presumably, at this point the target had been processed to the extent that subsequently presented stimuli exert little or no effect on the forthcoming manual response.

General discussion

In the present study we manipulated the level of competition between a target and a distractor stimulus by presenting them either in the same or in opposite visual fields, on opposite sides of the vertical meridian. Previous studies, both behavioural (Mounts & Gavett, 2004; Sereno & Kosslyn, 1991; Torralbo & Beck, 2008) and neurophysiological (Chelazzi et al., 1998; Sato, 1988, 1989), have shown that competitive interactions are strongest between stimuli presented in the same visual field. This is also what we observed: If a distractor and a target stimulus were presented simultaneously, RTs are about 35ms slower when both stimuli were presented in the same visual field, relative to when they were presented in opposite visual fields. The size of this RT difference is in the same range as previously reported in studies using similar measures (e.g., Sereno & Kosslyn, 1991).

Importantly, we have shown that in order for competition to arise, stimuli do not need to be presented at the exact same moment, although they do need to be presented in close temporal succession. In the present study, the 'window of competition' was asymmetrical, since we observed competition if the target was presented 50ms before the distractor, but little, if any, competition if the distractor was presented 50ms before the target. Tentatively, this can be ascribed to the fact that the distractor was task-irrelevant and therefore quickly inhibited (Theeuwes, Atchley, & Kramer, 2000), thus failing to compete with the subsequently presented target. The target, on the other hand, was task-relevant. Therefore,

it was presumably not inhibited and still strongly represented when the distractor was presented, resulting in competition.

Another important result was obtained when the distractor was presented prior to the target. As expected, we did not obtain evidence for direct competition between the target and the distractor at longer SOAs. However, we did obtain a striking pattern reversal: If the distractor was presented at least 125ms before the target, it caused more interference if the stimuli were presented in opposite visual fields, rather than in the same visual field. This finding illustrates that attentional effects depend strongly on whether stimuli are presented consecutively or simultaneously (Reynolds & Chelazzi, 2004). After a stimulus has captured attention, processing of another stimulus subsequently presented at the same location, is facilitated. Importantly, the strength of attentional facilitation is not strictly confined to the initial locus of attention, but rather falls of with increasing distance (Connor, Gallant, Preddie, & Van Essen, 1996; Connor, Preddie, Gallant, & Van Essen, 1997). This has been demonstrated in a number of behavioural studies which show that the locus of attention is not sharply delimited, but surrounded by a diffuse periphery which receives less, but still some attentional benefit (Downing, 1988; Downing, Pinker, Posner, & Marin, 1985; Mangun & Hillyard, 1988). Furthermore, it has been shown that this attentional benefit is greater for stimuli in the same visual field as the attended stimulus, relative to stimuli in the opposite hemifield (Hughes & Zimba, 1985, 1987; Rizzolatti et al., 1987; Tassinari, Aglioti, Chelazzi, Marzi, & Berlucchi, 1987). In the present study, the salient distractor presumably captured attention, even though it was task-irrelevant (Theeuwes, 1991). In order to make an accurate report of the target, attention had to be reoriented to the target. The fact that reorienting attention is costly gives rise to a distractor interference effect, measured in the present study by increased RTs relative to the control condition. However, it is known that reorienting attention to a new location within the same visual field is less costly than reorienting attention to a new location in the opposite visual field (Rizzolatti et al., 1987). In our data this is reflected by the fact that at longer SOAs distractor interference is greatest if the target and the distractor are presented in opposite visual fields.

In the present study we have used the term visual field to refer to the left and right visual hemifields, which are separated by the vertical meridian. However, a number of studies have shown that there is also a cost of crossing the horizontal meridian (Rizzolatti et al., 1987; Umiltà, Riggio, Dascola, & Rizzolatti, 1991). That is, if attention is allocated to some location, there is a cost of reorienting across the horizontal as well as the vertical meridian. In the present paradigm, we were unable to distinguish between these two

meridian effects: if the target and the distractor were presented on the same side of the vertical meridian, they were presented on opposite sides of the horizontal meridian and vice versa. However, for a number of reasons we feel that an explanation in terms of the vertical meridian is appropriate. Most importantly, behavioural studies on biased competition have reported an effect of the vertical meridian, but no effect of the horizontal meridian (e.g., Mounts & Gavett, 2004). Furthermore, there is a clear anatomical separation between the left and the right cerebral hemisphere and each of the two hemispheres deals predominantly with information from the contralateral visual field, at least in the early stages of visual processing (Desimone & Schein, 1987; Gattass et al., 1988). Even in areas of the brain where the separation between visual fields is not as strict, there is less competition between stimuli presented on opposite sides of the vertical meridian, but not the horizontal meridian (Chelazzi et al., 1998).

An important implication of the present study is that a lack of direct competition does not imply that the relationship between stimuli is neutral. The presentation of a stimulus may affect processing of another stimulus, presented up to half a second later. This is an important consideration when investigating biased competition: manipulating whether stimuli are presented simultaneously or sequentially (e.g., Beck & Kastner, 2005; Kastner et al., 1998) may yield results which are not solely due to the lack or presence of competitive interactions. Specifically, our results show that the effects of competitive interactions may in some cases appear larger than they really are, when operationalised as the difference between simultaneous and sequential presentation.

In summary, in the present study we investigated the temporal characteristics of distractor interference in the processing of a target stimulus. Specifically, we investigated how distractor interference varies as a function of whether a target and a distractor are presented in the same or in opposite visual fields. The most striking result is that the direction of this hemifield effects depends on whether stimuli are presented sequentially or simultaneously (see Figure viii.2c). By combining direct competition and sequential presentation in a single paradigm, we have provided a clear demonstration of the dynamic nature of visual attention. Biased competition explains why target processing is disrupted if a target and a distractor are presented (almost) simultaneously. However, if a target and a distractor are presented sequentially, interference reflects the cost of reorienting attention from the distractor to the target.

∞

PART 2: BEHAVIOURAL METHODS

IX. OPENSESAME: AN OPEN-SOURCE, GRAPHICAL EXPERIMENT BUILDER FOR THE SOCIAL SCIENCES

Abstract — In the present paper we introduce OpenSesame, a graphical experiment builder for the social sciences. OpenSesame is free, open-source, and cross-platform. It features a comprehensive and intuitive graphical user interface, and supports Python scripting for complex tasks. Additional functionality, such as support for eye trackers, input devices and video playback is available through plug-ins. OpenSesame can be used in combination with existing software for creating experiments.

Resources — OpenSesame is freely available from http://www.cogsci.nl/opensesame

∞

A little over 20 years ago, Schneider (1988) estimated that it took a professional programmer approximately 160 hours to implement a new experimental paradigm. Fortunately, things have changed since then. Advances in both software and hardware have made it possible for unskilled programmers to develop their experiments rapidly, using any of the available 'point-and-click' software packages (for a comparison, see Stahl, 2006). Scientists who prefer programming over the use of a graphical interface also benefit from the power of modern, high level programming languages, which have substantially improved the readability and reduced the amount of code required to perform most tasks.

Another important development is the increased availability of high quality, free experimental software. Currently, there are at least eight free software packages that are viable tools for creating experiments (see Table ix.1). However, these packages occupy a relatively small niche, mostly because they do not offer the type of fully graphical interface that many users have come to expect. Therefore, researchers that are most comfortable in a graphical environment generally use proprietary software.

Table ix.1. An overview of software for creating experiments.

Name	GUI	Free	Scripting	Platform	Reference/ Vendor
DirectRT	Yes	No	Custom	Windows	Reviewed in Stahl (2006)
DMDX	No	Yes*	Custom	Windows	K. I. Forster & J. C. Forster (2003)
E-Prime	Yes	No	E-Basic	Windows	Reviewed in Stahl (2006)
Experiment Builder	Yes	No	Python	Windows	SR Research, Missisauga, ON, Canada
Inquisit	Yes	No	Custom	Windows	Reviewed in Stahl (2006)
Matlab Psychophysics toolbox	No	Yes**	Matlab	Windows, Mac OS, Linux	Brainard (1997)
MEL	No***	No	Custom	IBM PC	Schneider (1988)
PEBL	No	Yes	Custom	Windows, Mac OS, Linux	Mueller (2010)
Presentation	Yes	No	Custom	Windows	Neurobehavioral Systems, Albany, CA, United States
PsychoPy	Yes	Yes	Python	Windows, Mac OS, Linux	Peirce (2007)
PsyScope	Yes	Yes	Custom	Mac OS	Cohen, MacWhinney, Flatt, & Provost (1993)
PsyToolkit	No	Yes	Custom	Linux	Stoet (2010)
PyEPL	No	Yes	Python	Mac OS, Linux	Geller, Schleifer, Sederberg, Jacobs, & Kahana (2007)
SuperLab	Yes	No	Custom	Windows	Reviewed in Stahl (2006)
Tscope	No	Yes	C/C++	Windows****	Stevens, Lammertyn, Verbruggen, & Vandierendonck (2006)
Vision Egg	No	Yes	Python	Windows, Mac OS, Linux	Straw (2008)

* Source-code is not available

** Depends on MATLAB (The MathWorks, 1998), a proprietary software package, for full functionality. Offers limited support for Octave (Eaton, 2002), an open-source MATLAB clone.

*** Uses a form-based interface

**** Offers limited support for Mac OS and Linux

In the present paper we introduce OpenSesame, a new experiment builder. OpenSesame is unique in that it is free, cross-platform, and arguably offers the most

intuitive and comprehensive graphical user interface (GUI) currently available. For complex tasks, which cannot be performed through the GUI, OpenSesame supports Python scripting (Van Rossum, 2008). A wide range of experiments can be created, including psychophysical experiments, speeded response time tasks, eye tracking studies and questionnaires.

In the first section of the present paper, we provide a non-technical description of the functionality offered by OpenSesame. In the second section, we describe how OpenSesame compares to, and can be used in combination with, existing software. The third section deals with timing considerations, and is followed by the fourth section in which the results of a benchmark experiment are described.

OpenSesame is freely available for download from http://www.cogsci.nl/opensesame. Documentation, a step-by-step tutorial, example experiments, and plug-ins can be found in the documentation area at http://osdoc.cogsci.nl/. The version of OpenSesame reviewed in the present paper is 0.24. At the time of writing, OpenSesame has been downloaded approximately than 10,000[14] times (non-unique downloads from cogsci.nl and via bit-torrent) and there is an active support forum.

Usage and functionality

System requirements

OpenSesame does not impose strict system requirements. Installation packages are available for Windows XP/ Vista/ 7 and Ubuntu/ Debian Linux. OpenSesame has been extensively tested on those platforms. At the time of writing, packages for Mac OS are also available, but have been labelled "experimental" pending further testing. On other platforms, users will need to manually install the software on which OpenSesame depends and run OpenSesame from the source code. Instructions for running OpenSesame from source are provided online.

The processing power required to run OpenSesame depends strongly on the type of experiment. For a typical experiment, consisting of a sequence of static stimulus displays followed by response collection, the requirements are very modest and any relatively modern computer system will suffice (we have successfully run OpenSesame on a low-end netbook; 2GB Ram, 1.66GHz Intel Atom N270). When using complex stimuli, such as high definition video, the user will need to evaluate for him or herself whether the computer system is up to the task.

---

14   As of May 24th 2012, the number of non-unique downloads exceeds 40,000.
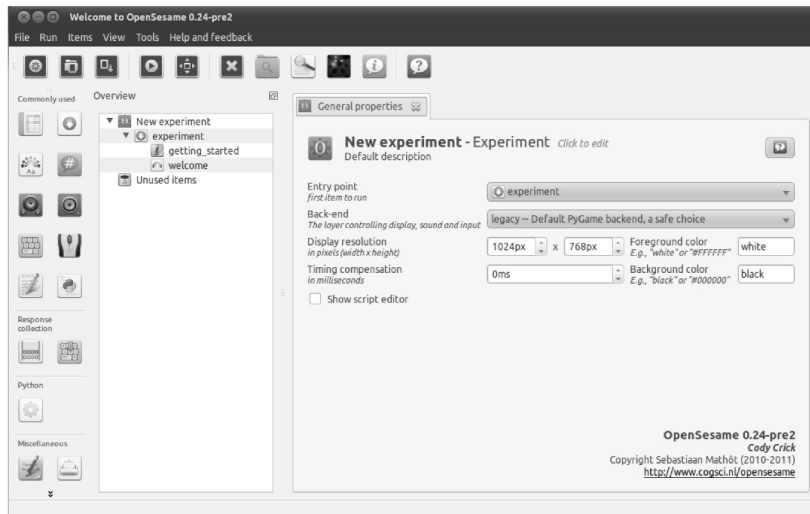
The graphical user interface

After starting OpenSesame, the user is presented with a GUI (see Figure ix.1). The structure of the experiment is depicted graphically as a tree-structure in the overview area (Figure ix.1b). An experiment consists of a collection of *items*, which are self-contained parts of the experiment (i.e., conceptually distinct units). Items can be added to the experiment by dragging them from the item toolbar (Figure ix.1a) onto the overview area. Items can be edited by selecting them in the overview area, after which the appropriate controls appear in the tab area (Figure ix.1c).

There are 10 core items that provide the basic functionality for creating an experiment (see Table ix.2 for an overview; functionality can be extended as described in *Usage and functionality: Plug-ins*). Examples of items are the *sketchpad*, which handles the presentation of a single stimulus display, and the *keyboard_response*, which handles the collection of a single keyboard press. Two special items are the *loop* and the *sequence*, which control the structure of the experiment. A *sequence* item sequentially calls a number of other items. A *loop* item repeatedly calls a single other item, while varying the values of independent variables. By combining items in various ways, arbitrary experimental paradigms can be constructed (Figure ix.1d).
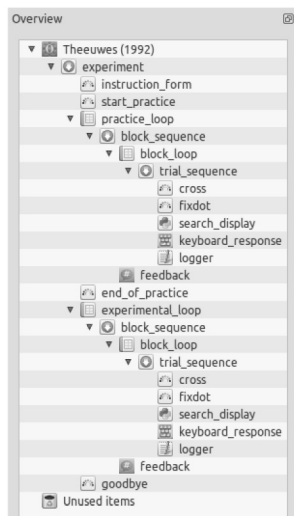
Table ix.2. An overview of the 10 core items

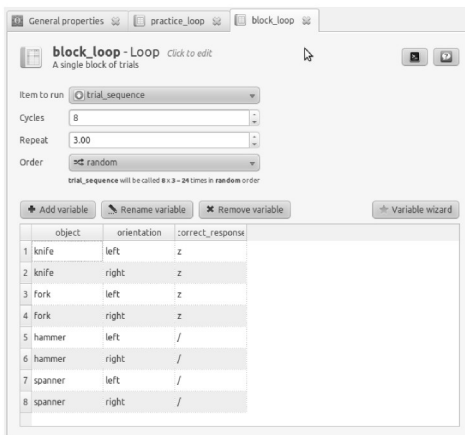| Name	Type	Description
loop	Structure	Repeatedly runs a single other item. Controls independent variables.
sequence	Structure	Runs multiple other items in sequences. Supports basic conditional statements ("Run if ...").
sketchpad	Stimulus presentation	Provides a canvas for presentation of visual stimuli.
feedback	Stimulus presentation	Provides feedback to participants.
sampler	Stimulus presentation	Plays a sound from file.
synth	Stimulus presentation	Provides basic sound synthesis.
keyboard_response	Response collection	Collects keyboard responses.
mouse_response	Response collection	Collects mouse responses.
logger	Data logging	Writes variables to file.
inline_script	Inline scripting	Executes arbitrary Python code.

a) Item
toolbar

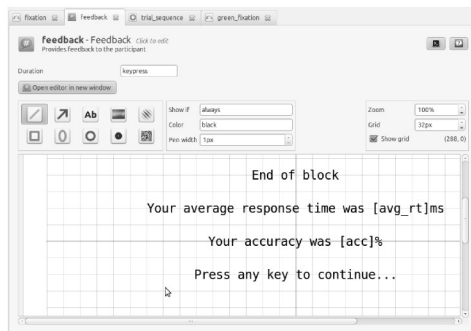b) Overview area

c) Tab area

d) Example experimental
structure

Figure ix.1. The OpenSesame graphical user interface on start-up. a) The item toolbar contains icons that can be dragged into the overview area. b) The overview area represents the experiment as a tree-structure. c) The tab area contains tabs for editing items and getting context sensitive help. By clicking on an item in the overview area, a tab with the appropriate controls opens. By clicking on one of the blue "help" buttons, a context sensitive help tab opens. d) The structure of an example experiment shown in the overview area.
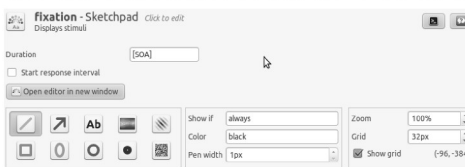
a) variables defined in a loop

c) using variables to provide feedback

b) using a variable to specify the duration of a sketchpad

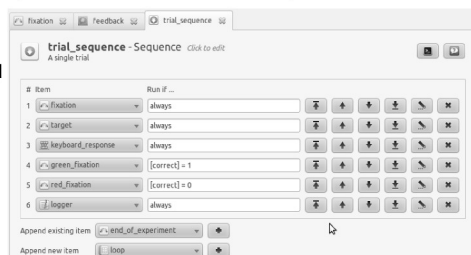d) conditional statements in a sequence

Figure ix.2. Using variables and conditional ("if") statements. a) Independent variables can be defined in a *loop* item. b) By entering *[SOA]* in the duration field of a *sketchpad* item, the variable *SOA* is used to control the presentation duration of the sketchpad. This assumes that *SOA* has been defined elsewhere in the experiment. c) By using *[avg_rt]* and *[acc]* as part of the text in a feedback item, appropriate feedback can be given to the user. The variables *avg_rt* and *acc* are set automatically by the various response items (e.g., *keyboard_response*). d) Conditional statements can be used to control which items from a *sequence* will be called. By entering *[correct] = 1* in the "Run if..." field of "green_fixation", the item will be called only if the variable *correct* has been set to *1*. The variable *correct* is set automatically by the various response items.

Variables and conditional ("if") statements

One of the biggest challenges when designing a GUI is to offer sufficient flexibility. In OpenSesame, this flexibility is achieved through the support of variables and conditional ("if") statements.

Variables can be built-in (e.g., *subject_nr*), set by items (e.g., *response_time*, which is set by *keyboard_response* items) or specified by the user in *loop* items (see Figure ix.2a). These variables can be used throughout the GUI, by entering *[variable_name]* in places where you would normally encounter a static value. For example, if the user has defined a variable called *SOA*, this variable can be used to control the duration of a *sketchpad* item (i.e., a stimulus display) by entering *[SOA]* in the duration field (Figure ix.2b).

Analogously, feedback of the average response times can be given by adding the text "Your average response time was [avg_rt]msg" to a *feedback* item (Figure ix.2c).

Particularly powerful is the possibility to combine the what-you-see-is-what-you-get *sketchpad* drawing tool (shown in Figure ix.2b,c) with the use of variables. The drawing tool automatically generates a simple script that defines the elements in the *sketchpad*. By replacing the static coordinates, colours, sizes, etc. with variables, the user can create a flexible stimulus display in an intuitive way. For example, if you insert an image (from the file *fork_left.png*) in the centre of the display, OpenSesame will generate the following line of script (taken from the *affordances_orientation* example experiment, available online):

```
draw image 0 0 "fork_left.png" scale=1 center=1 show_if="always"
```

By right-clicking on the object, you are given the possibility to edit this line. The static values can be replaced by variables, and thus the presented image as well as the image's size can be made variable (this will result in the object being hidden from the drawing tool, with the message that the sketchpad contains a variably defined object):

```
draw image 0 0 "[object]_[orientation].png" scale=[scale] center=1
show_if="always"
```

Conditional statements, commonly referred to as "if"-statements, can be used to add even more flexibility to the GUI. Every item from a *sequence* item has a "Run if" parameter that is evaluated before the item is called. This can be used, for example, to show a green or red fixation dot depending on whether the preceding response was correct (using the *correct* variable, which is automatically set by response items; Figure ix.2d). Analogously, in *sketchpad* and *feedback* items conditional statements can be used to control which elements are actually shown, by setting the "Show if" parameter.

Data output format

Usually, data logging will be handled by the *logger* item. Every time that a *logger* item is called, a single row of data is written to the output file. The output file is a plain-text comma-separated spreadsheet (.csv), in which each row typically corresponds to a trial, and each column corresponds to a variable. This format is compatible with all commonly used spreadsheet software.

Alternatively or in addition, users can write arbitrary messages to the output file using Python code in *inline_script* items. In this case, the format of the output file is determined by the user.

Python inline coding

Despite the flexibility of the GUI, there will sometimes be situations that require a full fledged programming language. For this reason, OpenSesame supports Python scripting
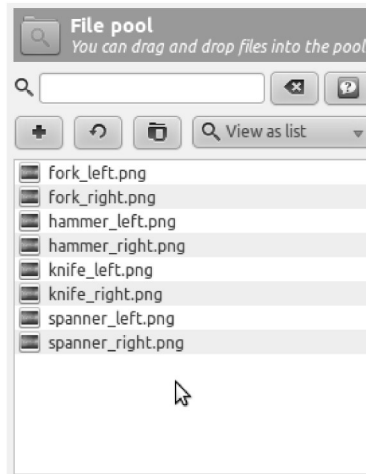
Figure ix.3. The file pool. All files that are used in the experiment, such as images, sounds and videos, are stored in the file pool. The file pool is saved along with the experiment, allowing for maximum portability.

(Van Rossum, 2008). Python is a powerful programming language that is widely used in the scientific community and the computing world in general (the 7th most widely used programming language, cf. Tiobe.com, 2011).

To use Python scripting, you add an *inline_script* item to the experiment. Rather than the knobs, buttons and input fields that you will see when editing other items, the *inline_script* item offers an embedded programming editor. You can use the OpenSesame Python modules, which offer simple routines for handling display presentation, sound generation, response collection, etc. Alternatively, you can interact directly with the selected back-end (see *Usage and functionality: The back-end layer*). This latter option is mostly useful for advanced users who are familiar with one of the back-ends (i.e., PsychoPy (Peirce, 2007), PyGame, and/ or OpenGL) or if the native OpenSesame modules do not offer the desired functionality.

File format and the file pool

External files, such as images, sound files and videos, are stored in the file pool (Figure ix.3). Optionally, OpenSesame also copies participant data files to the file pool after an experiment has finished. The file pool is saved along with the experiment in a single *.opensesame.tar.gz* file. This extension is somewhat ungainly, but accurately reflects that the file is a *.tar.gz* archive, which can be opened in most archiving tools (e.g., WinRar or FileRoller). Therefore, in the unlikely event that OpenSesame fails to open the experiment,

you can still access the experiment script and any files that have been stored in the file pool.

Plug-ins

OpenSesame can be extended through plug-ins. Plug-ins offer graphical controls, appear in the item toolbar and can be dragged into the experiment just as the built-in items. Therefore, from the perspective of the user there is little difference between using a plug-in and using any of the 10 core items.

Plug-ins offer arbitrary functionality, such as support for specific devices or handling particular tasks that would otherwise have to be implemented through Python inline code. Currently, there are plug-ins available that support the Eyelink series of eye-trackers (SR Research, Mississauga, ON, Canada), the Mantra object tracker (Mathôt & Theeuwes, 2011b), the Serial Response box (Psychology Software Tools, Sharpsburg, PA, USA), input devices connected to the parallel port, and video playback.

The back-end layer

OpenSesame consists of three distinct layers. The top layer handles the GUI and offers the functionality needed to create an experiment. The middle layer handles the execution of an experiment. The bottom, or back-end, layer handles the interaction with the display, sound and input devices.

There are many different Python libraries available that could, in principle, be used in the back-end layer. Some, such as PyGame, Pyglet, and PyOpenGL are general purpose, mostly oriented towards development of games. Others, such as PsychoPy (Peirce, 2007), VisionEgg (Straw, 2008) and PyEPL (Geller et al., 2007) have been developed specifically for creating psychological experiments.

OpenSesame is back-end independent, in the sense that different libraries can be used in the back-end layer. New back-ends can be created and added, much like plug-ins. Currently, there are three back-ends available: the *legacy* back-end, that uses PyGame, the *psycho* back-end, that uses PsychoPy (Peirce, 2007), and the *opengl* back-end that uses PyGame in combination with PyOpenGL (see Table ix.3).

Table ix.3. An overview of OpenSesame back-ends

Back-end name	Hardware accelerated	Underlying technology
legacy	No	PyGame in non-OpenGL mode
opengl	Yes	PyGame in OpenGL mode
psycho	Yes	PsychoPy in Pyglet window mode (Peirce, 2007)

From the perspective of the GUI user, there is little noticeable difference between the back-ends (although there may be small differences in anti-aliasing and font rendering). A red square will always be a red square, regardless of which library is used to draw it. Nevertheless, there are compelling reasons for having different back-ends to choose from. The first is that not all back-ends may be equally well supported on all platforms. Second, back-ends may differ in their timing properties (see *Benchmark experiment*). Third, each back-end offers unique functionality that the user can exploit when writing Python inline code. For example, if the *psycho* back-end is enabled, you can use the PsychoPy routines for creating visual stimuli. More generally, users can select the back-end that they are most familiar with and best suits their own approach to creating experiments.

Comparison to, and interoperability with existing software

Table ix.1 provides a list of the most popular software for creating experiments. There are already many software packages to choose from, yet the functionality offered by OpenSesame is unique. In this section, we will focus on two packages with which OpenSesame arguable has most in common: E-Prime (Psychology Software Tools, Sharpsburg, PA, United States) and PsychoPy (Peirce, 2007).

E-Prime (Psychology Software Tools, Sharpsburg, PA, United States) has been built on the legacy of the Micro Experimental Library (MEL; Schneider, 1988). Largely because it was one of the first programs to offer a graphical environment for creating experiments, E-Prime has, in our experience, become the de facto standard in many psychological labs around the world. There are at least four important differences between E-Prime and OpenSesame. The most obvious difference is that OpenSesame is free and open-source, whereas E-Prime is non-free and proprietary. Because of the open character of OpenSesame and the availability of a plug-in framework, it is easy for third parties to add and modify functionality. Second, OpenSesame is cross-platform, whereas E-Prime is exclusively available for Windows. A final, crucial difference is the language that is used for inline coding. Whereas E-Prime uses E-Basic, a dialect of the well-known BASIC language, OpenSesame uses Python (see *Usage and functionality: Python inline coding*).

The advantage of using Python over E-Basic is the availability of a large number of libraries, many of which are oriented towards the scientific community (e.g., Jones, Oliphant, & Peterson, 2001).

PsychoPy is an open-source project, which has gained considerable momentum as a comprehensive and well maintained Python library for creating psychological experiments (Peirce, 2007). Like OpenSesame, PsychoPy is cross-platform and open-source. And like OpenSesame, PsychoPy offers both a GUI (the 'builder view') and a set of Python libraries. However, there are substantial differences between the graphical interfaces offered by both packages. In the builder view of PsychoPy, the temporal progression of the experiment is shown, but the spatial arrangement of the stimuli is not readily apparent. In OpenSesame, the temporal progression is shown as well (in the overview area; Figure ix.1b,d), but in addition the user can get a visual preview of the stimulus arrangements through the *sketchpad* item (Figure ix.1c). Depending, of course, on the prior experience and personal preference of the user, such a preview can be very helpful. OpenSesame also offers a number of advanced features, which are not available in the PsychoPy builder view, such as integrated drawing tools, a more advanced sound synthesizer, and GUI support for widely used devices such as the serial response box (Psychology Software Tools, Sharpsburg, PA, USA) and the Eyelink series of eye trackers (SR Research, Mississauga, ON, Canada). More generally, OpenSesame and PsychoPy differ in their target audience and core functionality. Although PsychoPy provides a GUI, the focus is on the specialized set of Python libraries, which offer high level routines for creating complex visual stimuli, particularly those that are frequently used in psychophysical experiments (e.g, Gabor patches and random dot patterns). For this reason, PsychoPy will appeal mostly to people who prefer to code their experiments, such as MATLAB users who are looking for a viable open-source alternative. In contrast, OpenSesame offers only basic Python libraries, but has a comprehensive GUI, and will therefore appeal to users who are mostly at home in a graphical environment. As discussed previously (see *Usage: The back-end layer*), OpenSesame and PsychoPy can be used in combination. This allows users to get 'the best of both worlds', by combining the OpenSesame GUI with the PsychoPy Python libraries.

Timing

What is 'millisecond precision timing'?

A common and valid question that applies to all experiment building software is whether the timing is sufficiently precise. In our view, the best measure of a system's timing precision is the interval between the timestamp of a stimulus' presentation and the

timestamp of a response, given the fastest possible responder that is not directly part of the system itself. Put more simply, the lowest possible response time is a good indication of a system's timing precision. The reason for this is that any inaccuracies in the timing of display presentation and response collection will add to the lowest possible response time. Perhaps even more importantly, the lowest possible response time should be consistent, so that any delay is fixed and does not constitute a source of noise (see also *Benchmark experiment*).

However, even when a system has 'millisecond precision timing' in this sense, as most modern systems do, there are other factors that should be taken into account. First, many psychological labs, including our own, often use garden variety keyboards as input devices. Such keyboards have been designed to keep up with typing, and not to record responses with millisecond precision. As such, keyboards have been reported to have a relatively high and variable latency of up to 20ms (however, it is debatable whether this constitutes a significant source of noise, when considering the much larger variability in human response times; for a discussion, see Damian, 2010).

Computer monitors also have a number of peculiar properties. Rather than being refreshed instantaneously, monitors are refreshed line by line from the top down and, for each line, pixel by pixel from left to right. On CRT (non-flat screen) monitors, there is only a single active pixel at any given time, so pixels are 'fading out' most of the time (which is why you can observe a flickering at low refresh rates). On TFT (flat screen) monitors, pixels remain active, so a refresh resembles a 'flood fill' from top to bottom (which is why you don't observe flickering on a TFT monitor)[15]. Most researchers are aware that it is best to synchronize the presentation of a new display with the moment that the monitor starts refreshing from the top ('synchronization to vertical refresh' or 'v-sync'). However, even if this is done, displays do not appear instantaneously. With a refresh rate of 100Hz, a stimulus presented at the top of the display will appear up to 10ms before a stimulus presented at the bottom of the display. Another consequence of the monitor's refresh cycle is that displays cannot be presented at arbitrary points in time. Again, with a refresh rate of 100Hz, the interval between two stimulus displays is always a multiple of 10ms. Attempting to use different intervals will lead to dropped frames or aberrant timing.

The prepare-run strategy

Psychological experiments typically consist of short intervals, called 'trials', during which a participant perceives stimuli and performs a task. Timing should be controlled

---

15 A video that shows a side by side comparison of CRT and TFT displays, recorded using a high speed camera, can be found here: http://www.vimeo.com/smathot/monitordemo

during a trial, but some unpredictable variation in the duration of the interval between trials is acceptable (cf. Schneider, 1988). Therefore, the best strategy for experimental software is to perform time consuming tasks before a trial, and to keep the operations that need to be performed during a trial to a bare minimum.

OpenSesame implements this strategy by calling each element from a *sequence* item twice. During the *prepare* phase, items are given time to prepare, for example by generating a sound in the case of a *synth* item, or by creating an 'offline canvas' in the case of a *sketchpad*. During the *run* phase, items simply execute a very limited number of trivial functions, such as showing a previously constructed canvas. This strategy substantially reduces the risk of timing glitches. The prepare-run strategy is implemented at the level of *sequence* items, which will typically contain the time-critical parts of an experiment. If a *sequence* is called multiple times by a *loop*, the loop as a whole is not prepared. Doing so would quickly lead to excessive memory usage and potentially cause rather than prevent timing issues.

Testing your own system

The timing properties of OpenSesame, or any other experiment builder, depend on the computer set-up that is used. Therefore, OpenSesame comes with a basic test experiment (*test_suite*) that allows you to run a number of checks. Most importantly, the test experiment checks whether the reported interval between two presented *sketchpads* matches the specified interval. The same test allows you to verify that synchronization to the vertical refresh is enabled (see *Timing: What exactly is "millisecond precision timing"?*). If not, 'tearing', in the form of horizontal lines running through the display, will be readily apparent.

Benchmark experiment

Even though the included test experiment allows you to run some useful checks (see *Timing: Testing your own system*), it is important to note that this form of testing relies on the computer's self-report. And this can be misleading. Specifically, the computer may report that a display has been presented, when, in fact, it has only been queued and will be presented some time later. The aim of the present experiment was therefore to check timing precision more rigorously, and to provide a list of benchmarks that allow the reader to evaluate whether OpenSesame is sufficiently precise to be useful in his or her experimental setting.

We used an artificial responder that responds, for all intents and purposes, instantaneously to an increase in light intensity. We measured the response times to a white

display. Both inaccuracies in the reported time of display presentation (assuming that inaccuracies are always such that the display is presented after the reported time, and never before) and inaccuracies in the reported time of the response (assuming that inaccuracies are always such that the reported time of the response is after the actual response) add to the lowest possible response time. The lowest possible response time can therefore be used as a measure of the combined timing precision of display presentation and response collection.

Methods

Two different test systems were used (the system specifications are listed in Table ix.4). A black display was presented for 100ms, followed by a white display. Automated responses to the white display were collected using a modified button-box that responds to an increase in light intensity (i.e., the presentation of the white display) through a photo-resistor. This photo-resistor was attached to the top-left of the monitor. To ascertain that the we obtained realistic results (i.e., not tweaked to optimize performance), the test experiment was created entirely using the GUI and no Python inline code was used. The Serial Response Box plug-in was used to interface with the button-box.

The experiment was run under Windows XP (Systems 1 and 2) and Ubuntu Linux 10.04 (System 2). The following combinations of resolution and refresh rate were tested: 1680x1050 at 120Hz (System 1); 1280x1024 at 85Gz (System 2); 1024x768 at 60Hz (System 2). A colour depth of 32 bits was used in all cases. All three currently available back-ends were tested: *legacy*, *opengl* (Linux only), and *psycho* (see *Usage and functionality: The back-end layer*). 1000 responses were collected for each test. During the test, we checked visually for 'tearing', which indicates that the synchronization to the vertical refresh is not enabled.

Results and discussion

The results of the experiment are shown in Figure ix.4 and, in more detail, in Table ix.4. No further statistics were performed, as the data are essentially descriptive. Synchronization to the vertical refresh was enabled in all cases except for the *legacy* back-end on Ubuntu Linux 10.04. The results clearly show that for time critical experiments it is advisable to use one of the hardware accelerated back-ends, either *psycho* or *opengl*. With these, the response times are consistently below 4ms on Windows XP and 2ms on Ubuntu Linux 10.04. This negligible delay should be acceptable in even the most time critical experiments.

Figure ix.4. Results of the benchmark experiment. The automated response times are shown as a Tukey box plot conform Robbins (2004). The central line reflects the median value. The rectangle shows the interquartile range (the 25th percentile to the 75th percentile). The whiskers reflect the minimum and maximum values that fall within 1.5 times the interquartile range. Dots correspond to individual observations that fall outside of 1.5 times the interquartile range.

As an aside, the results do not show any evidence that display mode has an impact on performance (preparing a large canvas obviously takes more time, but this cost is hidden in the prepare phase, as discussed in *Timing: The prepare-run strategy*). This means that there is no reason to choose a low resolution, colour depth or refresh rate, unless much of the drawing operations are performed 'on the fly', rather than prepared beforehand.

Table ix.4. Results of the benchmark experiment

Operating system	Test system	Display mode	Back-end	Avg. RT (ms)	Std. RT (ms)	Sync to vertical refresh enabled
Windows XP	1 *	1680x1050 @ 120Hz	psycho	3.28	0.53	Yes
			legacy	11.94	2.45	Yes
	2 **	1024x768 @ 60Hz	psycho	3.18	0.53	Yes
			legacy	16.01	2.59	Yes
		1280x1024 @ 85Hz	psycho	3.02	0.53	Yes
			legacy	9.24	1.37	Yes
Ubuntu Linux 10.04 LTS ***		1024x768 @ 60Hz	psycho	1.56	0.41	Yes
			opengl	1.45	0.51	Yes
			legacy	3.59	4.30	No
		1280x1024 @ 85Hz	psycho	1.55	0.44	Yes
			opengl	1.87	0.55	Yes
			legacy	3.33	2.48	No

* Computer: Intel Core 2 DUO E8400, 3Ghz, 2Gb; Graphics adapter: ATI Radeon EA H4350 (discrete); Monitor: Samsung 2233RZ, TFT, 22"

** Computer: Intel Core 2 DUO E8400, 3 Ghz, 2Gb; Graphics adapter: Intel GMA 4500, Intel Q45/Q43 Express chipset (integrated); Monitor: Llyama vision master pro 454, CRT, 17"

*** Running Gnome 2.30 with the Metacity window manager. The compositing layer (i.e. "Compiz" or "Desktop Effects") was disabled.

Discussion

In the present paper, we have introduced OpenSesame, a graphical, open-source experiment builder for the social sciences. OpenSesame complements existing experiment building software in a number of ways.

First, OpenSesame offers the kind of fully graphical environment that, until now, was offered only by proprietary, non-free software.

Second, OpenSesame is extensible through plug-ins (see *Usage and functionality: Plug-ins*). Among other things, this means that support for external devices can be added and, once a plug-in has been created, this novel functionality will integrate seamlessly with the user interface. Currently available plug-ins offer support for the Eyelink series of eye trackers (SR Research, Mississauga, ON, Canada), the Mantra object tracker (Mathôt & Theeuwes, 2011b), the Serial Response Box (Psychology Software Tools, Sharpsburg, PA, USA), input devices connected to the parallel port, and video playback.

Third, OpenSesame supports Python scripting (see *Usage and functionality: Python inline coding*). Even though the aim of OpenSesame is to have a flexible user interface that allows you to create complex experimental paradigms in a graphical way, there will sometimes be occasions where there is a need for a full programming language. Python (Van Rossum, 2008) is a widely used language and has excellent support for scientific applications in general (Jones et al., 2001) and design of psychological experiments in particular (Geller et al., 2007; Peirce, 2007; Straw, 2008).

Fourth, OpenSesame aims for interoperability with existing software. Specifically, this means that OpenSesame can use different back-ends to handle display and input operations (see *Usage and functionality: The back-end layer*). If, for example, you want to use the PsychoPy (Peirce, 2007) routines for creating visual stimuli, you can select the psycho back-end. When this back-end is selected, all display and input operations will be handled by PsychoPy, and you will be able to use the PsychoPy routines in your experiment. Other back-ends can be created and added in much the same way as plug-ins.

Fifth, OpenSesame is cross-platform (see *Usage and functionality: System requirements*). This is particularly useful in environments where different operating systems are being used. For example, our own lab contains a mixture of Windows XP, Mac OS and Linux computers. With OpenSesame, experiments are fully portable between operating systems.

In summary, OpenSesame is a new graphical experiment builder for the social sciences. OpenSesame is unique in that it makes creating psychological experiments easy and accessible for everyone. And perhaps even fun.

<div align="center">∞</div>

X. MANTRA: AN OPEN METHOD FOR OBJECT AND MOVEMENT TRACKING

Abstract — Mantra is a free and open-source software package for object-tracking. It is specifically designed to be used as a tool for response collection in psychological experiments and requires only a computer and a camera (a webcam is sufficient). Mantra is compatible with widely used software for creating psychological experiments. In Experiments 1 and 2 we validated the spatial and temporal precision of Mantra in realistic experimental settings. In Experiments 3 and 4 we validated the spatial precision and accuracy of Mantra more rigorously by tracking respectively a computer controlled physical stimulus and stimuli presented on a computer screen.

∞

Object-tracking is a powerful method of response collection. There are many examples of studies which have addressed important questions by using object-tracking. For example, in a study by Tipper and colleagues (Tipper, Howard, & Jackson, 1997) participants reached for a target stimulus (a wooden block) while the position of their hand was tracked. In addition to the target, a distractor stimulus was presented. The crucial finding was that the reaching trajectory of the hand systematically veered away from the distractor. The authors interpreted this as evidence for competition between the target and the distractor, which is resolved by inhibiting the distractor location. Another example that illustrates the usefulness of object tracking is a study by Brenner and Smeets (1996). In their study, participants picked up a target stimulus (a brass disk), while their thumb and index-finger were tracked in order to measure hand-opening. The apparent size of the target was manipulated by presenting it among converging lines in various configurations. The crucial finding was that this perceptual illusion did affect the participants' judgment of the size of the target, but did not affect how wide they opened their hand to reach for the target. Brenner and Smeets (1996) interpreted this finding as evidence for separate visual

streams for action and perception (Goodale & Milner, 1992)[16]. Both the study by Tipper and colleagues (1997) and the study by Brenner and Smeets (1996) illustrate clearly that object-tracking is a unique and flexible tool, which allows researchers to investigate issues that cannot be investigated otherwise.

Even in situations where the use of a keyboard may be considered an adequate form of response collection, object-tracking can provide additional information. For example, keyboard presses are often used to investigate whether responses are faster (or slower) in one condition compared with another condition. This approach has a rich history and forms the basis of many classic psychological paradigms (e.g., Donders, 1969; Posner, 1980). However, some questions are difficult to answer based on response time alone. For example, is there a difference in the time of movement-onset or is there a difference in the velocity of the movement? Both possibilities could lead to a decrease in response time as measured using a keyboard. This question, and many others, can easily be investigated by tracking the location of a participant's hand throughout a trial.

Despite the obvious advantages of object-tracking as a method of response collection, object-tracking systems are used sparingly by experimental psychologists. The reason is that the required equipment is generally expensive and is not part of the 'default set of equipment' found in most psychological laboratories. In the present paper we introduce Mantra, a system for object-tracking, which has three crucial advantages. First, Mantra is released under an open-source license and is available free of charge. Second, Mantra requires only a computer and a camera (an ordinary webcam is sufficient). Therefore, Mantra allows object-tracking with general purpose, widely available equipment. Third, Mantra is designed specifically as a tool for experimental psychology. Therefore it integrates painlessly with software for creating psychological experiments, such as E-Prime (Psychology Software Tools, Sharpsburg, PA, USA), PsychoPy (Peirce, 2007), PyEPL (Geller et al., 2007) and OpenSesame (Mathôt, Schreij, et al., 2012). Compared with systems such as the Liberty tracking system (Polhemus), TrakSTAR (Ascension Technology Corporation) or the Optotrak System (Northern Digital), Mantra offers basic functionality. However, for many purposes, such as the study by Tipper and colleagues (1997) described above, this basic functionality is precisely what is needed.

---

16 In retrospect, this statement is inaccurate. Although it is true that optical illusions have been interpreted as support for the existence of two distinct visual streams by some authors, Brenner and Smeets (1996) explicitly reject this idea, and propose that illusions selectively affect the perception of particular spatial attributes. For example, an illusion may affect perceived motion, but not position (for a review, see Smeets, Brenner, de Grave, & Cuijpers, 2002).

The first section of the present paper provides a brief, non-technical description of Mantra. The following sections describe four experiments. Experiment 1 is a replication of the Müller-Lyer illusion (Müller-Lyer, 1889), which we have designed to validate the spatial precision of Mantra in a realistic experimental setting. Experiment 2 is a variant of the additional singleton paradigm (cf. Theeuwes, 1994), which we have designed to validate the temporal precision of Mantra, also in a realistic experimental setting. In Experiments 3 and 4 we investigated the spatial precision and accuracy of Mantra more rigorously, by tracking respectively a computer controlled physical stimulus and stimuli presented on a computer display. A detailed description of Mantra, installation packages, source-code and experimental data can be downloaded from http://www.cogsci.nl/mantra.

Usage

System requirements

Mantra is available as an open-source software package for Linux and integrates directly with experiments created in E-Prime (Windows XP) and Python (cross-platform). Mantra will run on any modern computer system, including low-end systems, such as the Intel Atom-based netbook used in Experiment 3. A camera (e.g., a webcam) is required.

Defining objects

The first step in using Mantra is to define one or more objects. Object-definitions are based on colour, which provides a robust and computationally cheap way to track multiple objects simultaneously and unambiguously. Therefore, it is important to use distinctly coloured objects. Stickers or coloured pieces of paper can be attached to objects that do not have a distinct colour themselves. The number of objects that can be tracked simultaneously is determined by the number of colours that are sufficiently distinct. In turn, this depends on factors such as lighting and camera settings. In practice it is feasible to track up to five objects (Figure x.1c). In order to define an object you simply hold it in front of the camera and select it in the object-definition window (Figure x.1b). The colour of the selected pixel is taken as the object-defining colour. The object now turns green, whereas the rest of the image turns red. This allows you to determine visually if the object is reliably detected and is not confused with other objects. By default Mantra compensates for luminosity, by representing colour values relative to luminosity (e.g., $R_{rel} = R - (R + G + B)/3$). Therefore, detection remains reliable even if luminosity varies: A red object that has been defined in the light is also detected in the shade.
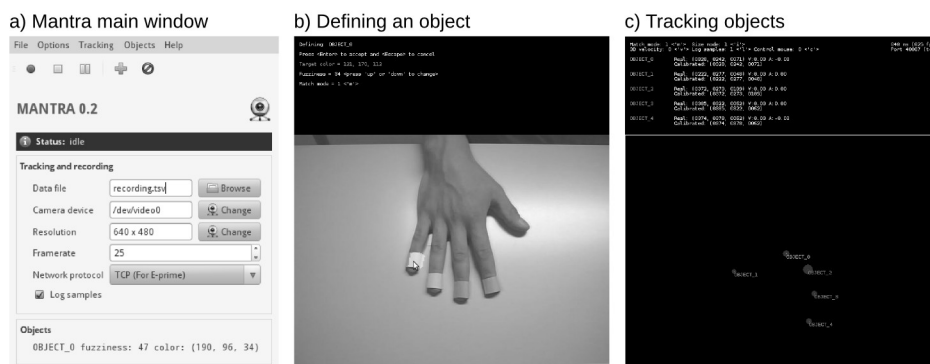
Figure x.1. Screenshots of Mantra. a) The main window contains controls to define objects, start tracking, change camera settings, etc. b) To define an object you click on it in the object-definition window. The selected object turns green whereas the rest of the image turns red. This allows you to check if an object is reliably detected before you start tracking. c) During tracking you can monitor the position of all objects. In this example the five fingers of a hand are being tracked.

Tracking

After all objects have been defined you can start tracking. While tracking is in progress you can monitor the location of the objects (Figure x.1c). The average location of all matching pixels is taken as the object's location (x, y). A z-coordinate is also available, which is defined as the maximum of the width and height of the object and can be used as a (very) coarse approximation of distance. The velocity and acceleration of the object are determined as well. If the velocity exceeds a certain threshold a movement start is signalled. If the velocity then drops below a second threshold a movement end is signalled. All data is logged as plain-text to a file.

In most cases the temporal resolution will be limited by the frame rate of the camera. Most webcams, including the webcams that we have used in our experiments, have a frame rate of 25Hz, which is equivalent to a temporal resolution of 40ms. On a 1.66GHz netbook, tracking at 25Hz, CPU consumption is around 53%, irrespective of the number of objects that are tracked (1 object: 53.1%; 5 objects: 53.7%).

The spatial resolution depends on two factors. The first factor is the resolution of the camera. In our experiments we have used a camera with a resolution of 640x480px, which is a typical resolution for webcams. The second factor is the distance between the camera and the object. For obvious reasons, spatial resolution is highest for objects near the camera. There is always a small jitter due to ambiguities in the separation between object and background (Figure x.3b; Figure x.7b,c,d). Under good conditions (i.e., with proper lighting, well defined objects, and using a camera with a resolution of 640x480 px),

objects can be tracked with a spatial precision of up to 0.3° (corresponding to about 2 mm in a regular set-up; see Experiment 3). Under optimal conditions (such as tracking ideal stimuli on a computer display) a measurement error of less than 0.1° is even feasible (Experiment 4).

Communication

Because Mantra is primarily intended as a data-collection tool for experiments, communication between the experiment and Mantra is crucial. Example code is provided in Listings x.1 (E-Basic) and x.2 (Python). The first step is to establish a connection between the experiment and Mantra. In order to do this you need to know the IP-address of the computer running Mantra, which depends on your network configuration. You must also know the port on which Mantra is listening, which is displayed in the tracking preview window (Figure x.1c). After a connection has been established the experiment can send information to Mantra. For example, the experiment can write messages to the Mantra log-file to indicate the start and end of a trial. The experiment can also retrieve information from Mantra. The coordinates of an object can be queried (Experiment 1) or the experiment can wait for the start or end of a movement (Experiment 2).

Listing x.1. Example E-Prime code. This example assumes that the E-Basic Mantra script has been included in the User Scripts section of the experiment.

```
1. ' Connect to Mantra assuming that Mantra is
2. ' running locally (IP-addres 127.0.0.1)
3. ' and listening on port 40007
4. If MConnect("127.0.0.1", 40007) = False Then
5.    ' Give a message on connection error
6.    Debug.print "Failed to connect to Mantra!"
7. Else
8.    ' Write to the Mantra log
9.    MLog "Waiting for movement!"
10.        ' Wait for object 0 to start moving
11.        MSMov 0
12.        ' Write to the Mantra log
13.        MLog "Movement detected!"
14.      End If
```

Listing x.2. Example Python code.

```
1. # Import the Mantra library
2. import libmantra
```

```
3.  # Connect to Mantra assuming that Mantra is
4.  # running locally (IP-address 127.0.0.1)
5.  # and listening on port 40007
6.  mantra = libmantra.libmantra("127.0.0.1", 40007)
7.  # Give a message on connection error
8.  if not mantra.connected:
9.    print "Failed to connect to Mantra!"
10. else:
11.   # Write to the Mantra log
12.   mantra.log("Waiting for movement!")
13.   # Wait for object 0 to start moving
14.   mantra.smov(0)
15.   # Write to the Mantra log
16.   mantra.log("Movement detected!")
```

Experiment 1

The first aim of Experiment 1 was to validate the spatial precision of Mantra in a realistic experimental setting. To this end we set out to replicate the Müller-Lyer illusion. The Müller-Lyer illusion refers to the fact that people tend to overestimate the length of a line-segment surrounded by inwards pointing arrowheads, relative to a line-segment surrounded by outwards pointing arrowheads (Müller-Lyer, 1889). In our experiment, participants controlled the length of a target line-segment by adjusting the distance between their thumb and index-finger, which were tracked by Mantra. A replication of the Müller-Lyer illusion in this way would be a compelling demonstration of the spatial precision of the Mantra system.

The second aim of Experiment 1 was to provide a demonstration of how Mantra can be used in combination with E-Prime (Psychology Software Tools, Sharpsburg, PA, USA). Because E-Prime is a widely used package for creating psychological experiments it is crucial that Mantra integrates well with E-Prime.

Method

Participants, stimuli and procedure

Five naive observers and one of the authors (SM) participated in the experiment (age range 18-27). All participants reported normal or corrected vision. The experiment was conducted in a well-lit room.

Before the start of each trial a grey fixation dot was presented on a black background for 500ms (Figure x.2a), followed by the presentation of two line-segments, 4.2° above and below the fixation dot. One of the line-segments was surrounded by inwards pointing
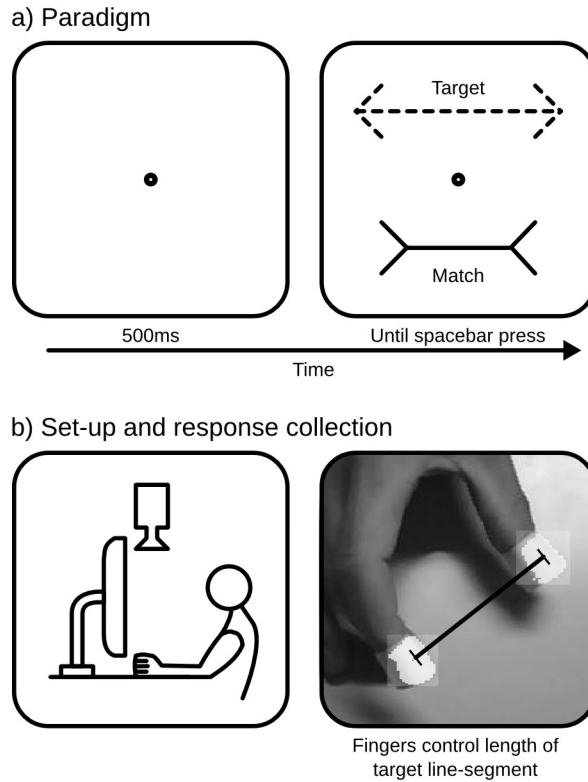
a) Paradigm

b) Set-up and response collection

Fingers control length of
target line-segment

Figure x.2. Paradigm, set-up and response collection in Experiment 1. a) Participants matched the length of the target line-segment (green) to that of the match line-segment and pressed the spacebar when they felt that both line-segments were equally long. b) The camera pointed downwards. Participants rested their right hand on the table surface underneath the camera. Participants manipulated the length of the target line-segment by changing the distance between their thumb and index-finger, which were tracked using Mantra.

arrowheads, the other line-segment was surrounded by outwards pointing arrowheads. One of the line-segments (the match) was grey and had a fixed length (a random value between 2.5° and 4.2°). The other line-segment (the target) was green and its length was adjusted online, based on the distance between the thumb and index-finger of the participant (see Apparatus, software and response collection). The arrowheads consisted of two lines, 1.7° in length. The arrowhead-style of the target (inwards-target/ outwards-match or outwards-target/ inwards-match) and the location of the target (target-above/ match-below or target-below/ match-above) were fully randomized. Participants were instructed to adjust the length of the target line-segment and to press the spacebar when they felt that both line-segments were equally long. It was emphasised that response time was not

important. The experiment consisted of 16 practice trials, followed by 128 experimental trials.

Apparatus, software and response collection

The experiment was run on a desktop computer (Intel Core Duo, 3 GHz, Windows XP) running E-Prime 1.2. Mantra 0.2 was run on a laptop running Linux (Intel Pentium T4300, 2.1 GHz, Ubuntu 9.10). Both computers were connected through an ethernet cable. For image acquisition a Logitech webcam was used, with a frame rate of 25Hz and a resolution of 640x480px. The webcam was mounted on top of the experimental display and pointed downwards (Figure x.2b). Participants wore a green paper 'fingercap' on their thumb and an orange fingercap on their index finger. The length of the target line-segment on the display (in display-pixels) was adjusted online to twice the distance (in webcam-pixels) between the thumb and index-finger.

Results

'Target-length' was defined as the length of the target line-segment relative to the match line-segment. Trials in which Target-length was less than 50% or more than 150% were excluded (0.1%). In total 99.9% of the trials were included in the analysis.

A two-tailed paired samples *t*-test revealed that Target-length was larger in the target-outwards/ match-inwards condition (M = 105.4%; SE = 1.3) than in the target-inwards/ match-outwards condition (M = 96.8%, SE = 1.7; $t(5)$ = 3.0; p < .05; Figure x.3a). All participants showed this effect, which reflects the Müller-Lyer illusion.

Figure x.3b shows Target-length over time for a single, representative trial. A number of things are apparent from this graph. First, the oscillations reflect the typical tendency to iteratively adjust, overshoot, and re-adjust the length of the target line-segment. Second, and more importantly, jitter due to measurement error is small. For example, during the first 400ms of this particular trial (the 10 frames before start of the first oscillation) the Target-length standard deviation is 0.4%.

Discussion

In Experiment 1 we replicated the Müller-Lyer illusion (Müller-Lyer, 1889). Participants controlled the length of a target line-segment by adjusting the distance between thumb and index-finger. The thumb and index-finger were tracked on a computer running Mantra and communicated to a second computer running the experiment (programmed in E-Prime), which dynamically adjusted the length of the target line-segment on the display.
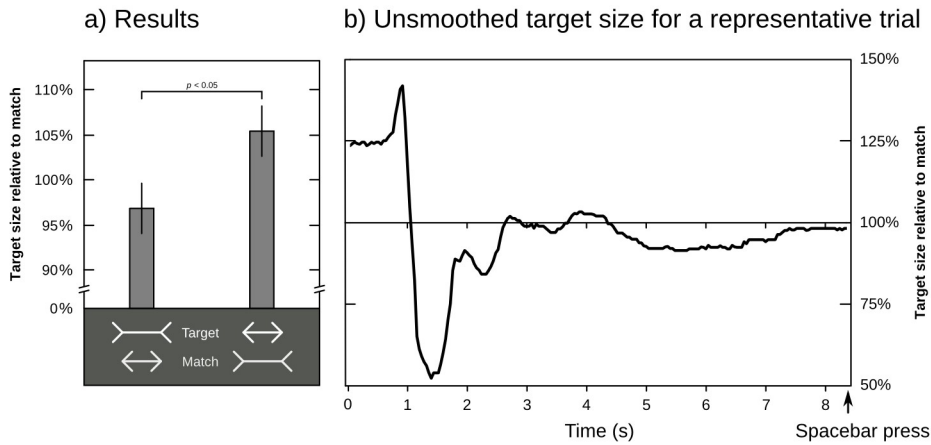
Figure x.3. The results of Experiment 1. a) Participants overestimated the length of the target line-segment when it was surrounded by inwards pointing arrows (and subsequently underadjusted its length), relative to when the target line-segment was surrounded by outwards-pointing arrows. This effect reflects the Müller-Lyer illusion (Müller-Lyer, 1889). Error-bars represent 95% within-subject confidence intervals (Cousineau, 2005). b) Unsmoothed target length relative to the match for a single, representative trial. This example shows the typical, slow oscillations in the size of the target line-segment while the participant is attempting to match its length. In addition, this example show that there is little jitter due to noise (the small, rapid oscillations).

Since it is conceivable that colour affects perceived size, a potential concern is that the target line-segment was always green, whereas the match line-segment was always grey. However, this would lead to a systematic over- or underestimation of the size of the target line-segment relative to the match line-segment, and cannot account for the Müller-Lyer illusion in the present experiment.

Two important conclusions can be drawn. First, Experiment 1 clearly shows that the position of multiple objects can be tracked reliably and precisely using Mantra. Second, Experiment 1 shows that Mantra integrates well with E-Prime.

Experiment 2

The first aim of Experiment 2 was to validate the temporal precision of Mantra. To this end we created a variant of the additional singleton paradigm, in which participants made a speeded report of the orientation of a line-segment within a uniquely shaped placeholder. Based on the literature, we expected that the presence of a distractor would result in increased response times, due to attention being captured by the distractor (Theeuwes, 1994). In addition, we expected that this distractor interference effect would be largest if the distractor was presented near the target, due to increased competitive interactions

between target and distractor at close spatial separations (Mathôt et al., 2010; Mounts, 2000b). In one condition, participants moved their index-finger, which was tracked by Mantra, to the left or to the right to make a response. In order to directly compare Mantra responses to keypress responses we also included a condition in which participants responded using a keyboard.

The second aim of Experiment 2 was to demonstrate how Mantra can be used in combination with Python. Interoperability with Python ensures that the use of Mantra does not require access to proprietary software. A number of packages are available for creating psychological experiments in Python, such as PsychoPy (Peirce, 2007), PyEPL (Geller et al., 2007) and OpenSesame (Mathôt, Schreij, et al., 2012).

Method

Participants, stimuli and procedure

Three naive observers and one of the authors (S.M.) participated in the experiment (age range 25-38). All participants reported normal or corrected vision. The experiment was conducted in a well-lit room.

Before the start of each trial a fixation dot was presented for 600ms (Figure x.4a), followed by the presentation of 6 premasks (size = 3.3°), arranged in a circle ($r = 10$°) centred around the fixation dot. All stimuli were grey and presented on a black background. Premasks consisted of a placeholder (a circle or a square, fully randomized) containing 6 line-segments, titled 0°, 30°, 60°, 90°, 120° and 150°. After 600ms all but one of the line-segments in each placeholder disappeared and one of the placeholders changed in shape (from a circle to a square or vice versa). Participants reported the orientation of the line-segment in the uniquely shaped placeholder. The target-line segment was always oriented horizontally or vertically. Non-target line-segments were never oriented vertically or horizontally. In 66% of the trials a distractor (identical to the regular placeholders) appeared simultaneously with the presentation of the target, at a random location midway between two regular placeholders. Distractor presence (absent/ present) was fully randomized. After a correct or incorrect response the fixation dot turned green or red, respectively.

The experiment consisted of two blocks (counterbalanced) which differed in response method. In the 'Keypress' condition participants pressed the 'Z'-key to report a horizontal line-segment and the '/'-key to report a vertical line-segment. In the 'Mantra' condition participants responded by moving their index-finger (see Apparatus, software and response
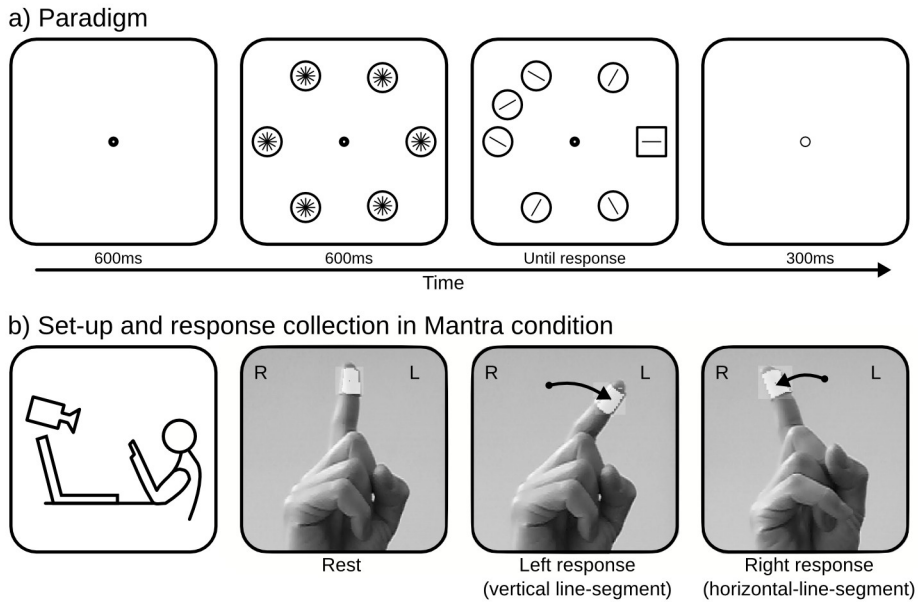
Figure x.4. Paradigm, set-up and response collection in Experiment 2. a) Participants reported the orientation of the line-segment in the uniquely shaped placeholder. In the Keypress condition participants responded using a keyboard. b) In the Mantra condition participants responded by moving their index-finger, which was tracked using Mantra, quickly to the left or to the right. The camera pointed slightly downwards towards the hand of the participant. Participants sat with their elbow resting on the table surface.

collection). In total, the experiment consisted of 40 practice trials and 480 experimental trials.

Apparatus, software and response collection

The experiment was run on a laptop running Linux (Intel Pentium T4300, 2.1 Ghz, Ubuntu 9.10) and written in Python. Mantra 0.2 was run on the same laptop. For image acquisition a Logitech webcam was used, with a frame rate of 25Hz and a resolution of 640x480px. The webcam was mounted on top of the laptop-display and pointed towards the participant in a slightly downwards angle (Figure x.4b).

In the Mantra condition participants wore a brightly coloured paper fingercap on their index-finger. To report a vertical line-segment, participants made a rapid left-wards movement with their index-finger. To report a horizontal line-segment, participants made a right-wards movement. For feedback purposes, movements were detected online using the standard Mantra movement detection algorithm. For the analysis, Reaction Time (RT) was

determined offline using an interpolation script (included in Mantra), which estimates the exact moment at which a movement is initiated.

Results

Trials in which RT was below 100ms or above 1200ms were excluded (2.5%). In total 97.5% of the trials were included in the analysis.

Separate, but identical analyses were performed for the Keypress and Mantra conditions. In the Keypress condition, a two-tailed paired samples *t*-test revealed that RTs were higher in the distractor present condition (M = 609ms, SE = 46.1), relative to the distractor absent condition (M = 564ms, SE = 49.6; $t(3)$ = 3.5, p < .05; Figure x.5a). In addition, for distractor present trials a repeated measures analysis of variance (ANOVA) was performed, using RT as dependent variable and distractor-target separation (Small, Medium or Large) as independent variable. There was a main effect of distractor-target separation ($F(2, 6)$ = 14.1, p < .01; Figure x.5b), such that there was more distractor interference if the distractor was close to the target.

The analysis for the Mantra-condition revealed qualitatively identical results. A two-tailed paired-samples *t*-tests with movement-onset time as dependent variable revealed an effect of distractor presence (present: M = 662ms, SE = 33.4; absent: M = 634ms, SE = 34.4; $t(3)$ = 7.1, p < .01). A repeated measures ANOVA with movement-onset time as dependent variable revealed a main effect of distractor-target separation ($F(2, 6)$ = 10.7, p < .05). All participants showed an effect of distractor presence as well as an effect of distractor-target separation, in both the Mantra and the Keypress-condition. A two-tailed paired samples *t*-test with maximum movement velocity as a dependent variable revealed no effect of distractor presence (present: M = 1.8px/ms, SE = 0.34; absent M = 1.8px/ms, SE = 0.32; $t(3)$ = 0.13, p = 0.9).

In order to investigate whether the level of noise was higher in the Mantra condition than in the Keypress condition, we calculated the standard deviation of correct RTs in distractor absent trials for each participant (using movement-onset time in the Mantra condition as the measure of RT). A paired samples *t*-test revealed no difference in RT standard deviation between the Keypress (M = 102.9) and Mantra condition (M = 100.0; $t(3)$ = 0.1, p = .9) in the distractor absent trials. The difference in overall RT between the Keypress (M = 593ms) and Mantra conditions (M = 653ms) did not reach significance ($t(3)$ = 3.2, p = .28).
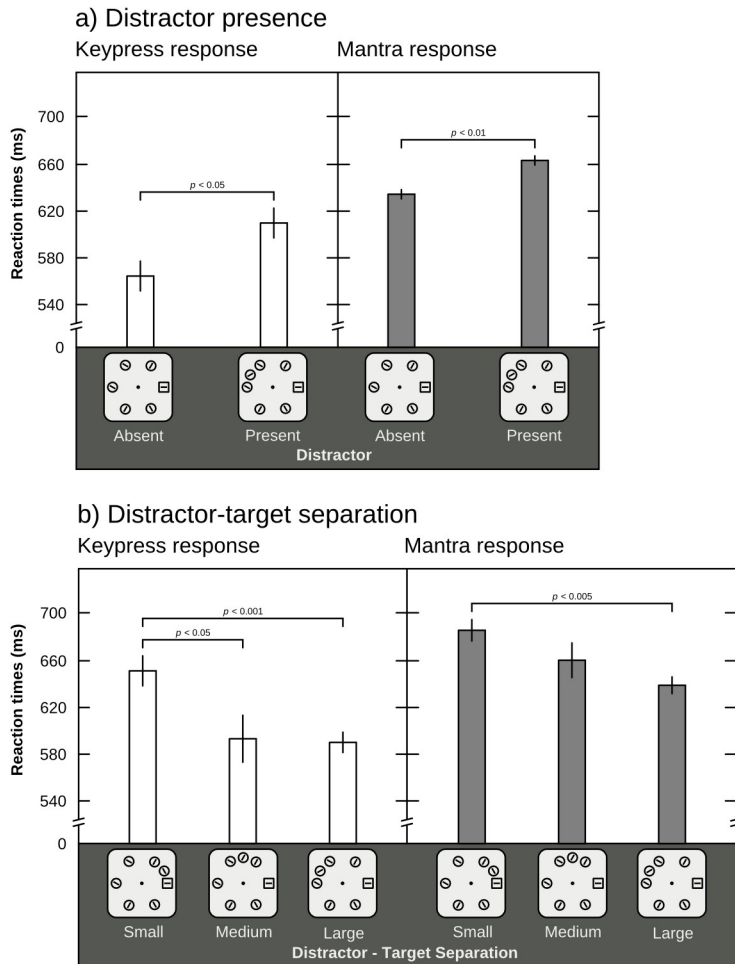
Figure x.5. Mean reaction times as a function of distractor presence (a) and distractor-target separation (b) in Experiment 2. Participants are slowed by the presence of a distractor (a; Theeuwes, 1994) and this effect is strongest if the distractor-target separation is small (Mathôt et al., 2010; Mounts, 2000). Importantly, the results in the Keypress and Mantra-conditions are qualitatively identical. Error-bars represent 95% within-subject confidence intervals (Cousineau, 2005).

Discussion

In Experiment 2 we replicated two typical findings from an additional singleton paradigm. Participants were slowed by the presence of a distractor (Theeuwes, 1994) and this effect was more pronounced if the distractor was presented near the target (Mathôt et al., 2010; Mounts, 2000b). Participants reported the orientation of a target line-segment

either by a keypress or by moving their index-finger in front of a camera. Both methods of response yielded quantitatively similar results. One difference is that the Keypress data suggest that distractor interference is essentially equal for medium and large distractor-target separations. In contrast, the Mantra data suggest that distractor interference decreases gradually as a function of distractor-target separation, which is actually more in line with theory and findings on biased competition (Mounts, 2000b).

Three important conclusions can be drawn from Experiment 2. First, the results clearly show that speeded responses can be registered precisely using Mantra. This may be surprising given the fact that responses were collected using a webcam with a frame rate of only 25Hz. Therefore, on any given trial there was a maximum temporal resolution of 40ms. However, this limited temporal resolution is only one of many sources of noise which contribute to the observed response times and are averaged out in the mean RTs (Brand & Bradley, 2011; Damian, 2010; Ulrich & Giray, 1989). Empirically, we have shown that Mantra is a viable tool for collecting speeded responses, even when used in combination with a camera with a limited temporal resolution.

Second, Experiment 2 shows that Mantra can be used to address questions which cannot be resolved using keypress responses. Specifically, we have shown that the presence of a distractor delays the initiation of a movement, but does not interfere (or very little) with speed of movement.

Third, we have shown that Mantra can be used in combination with Python. This is crucial, because Python is an open-source and platform-independent language, which makes it possible to use Mantra on different platforms and without the need for proprietary software.

Experiment 3

Experiments 1 and 2 showed that Mantra is a viable tool for response collection by demonstrating its use in realistic experimental settings. However, these experiments did not provide quantitative data on the precision of Mantra's measurements. Therefore, the aim of Experiment 3 was to quantify the precision with which Mantra is able to track a moving stimulus. We tracked a single stimulus that was attached via a mechanical arm to a computer controlled wheel.
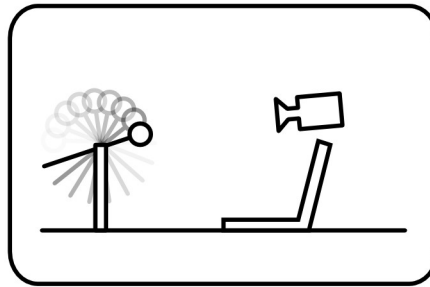
Figure x.6. Experimental set-up in Experiment 3. A stimulus, which was tracked by Mantra, was attached to a computer controlled rotating arm. In the actual set-up, the camera was positioned in front of the stimulus' rotational plane.

Method

Stimuli, apparatus and procedure

We attached a small orange sticker to the end of a mechanical arm, which was attached to a computer controlled wheel (see Figure x.6). Mantra ran on a netbook (Intel Atom, 1.66 GHz, Ubuntu 10.10) placed in front of the arm at 40 cm distance. The built-in webcam, with a spatial resolution of 640x480 px and a temporal resolution of 40 ms, was used for tracking.

The stimulus rotated in a continuous clockwise movement, describing a circular motion, with a radius of 9 cm, around the computer controlled wheel. We increased the speed of the stimulus in 9 steps from 10.7 cm/s to 52.8 cm/s.

Data analysis

Although we did not know the true position of the stimulus at any given time, we knew that the stimulus described a circular motion. Therefore, we judged the measurement precision by quantifying how well the measured trajectory resembled a circular motion. More specifically, we knew that the X- and Y-coordinates were described by sinusoidal functions. The measurement error was defined as the average absolute difference between the measured position of the stimulus, and the position of the stimulus as predicted by the two (i.e., one for the X-coordinate and one for the Y-coordinate) best fitting sines. This analysis was performed separately for each speed level. No further statistics were performed, as the data is essentially descriptive.
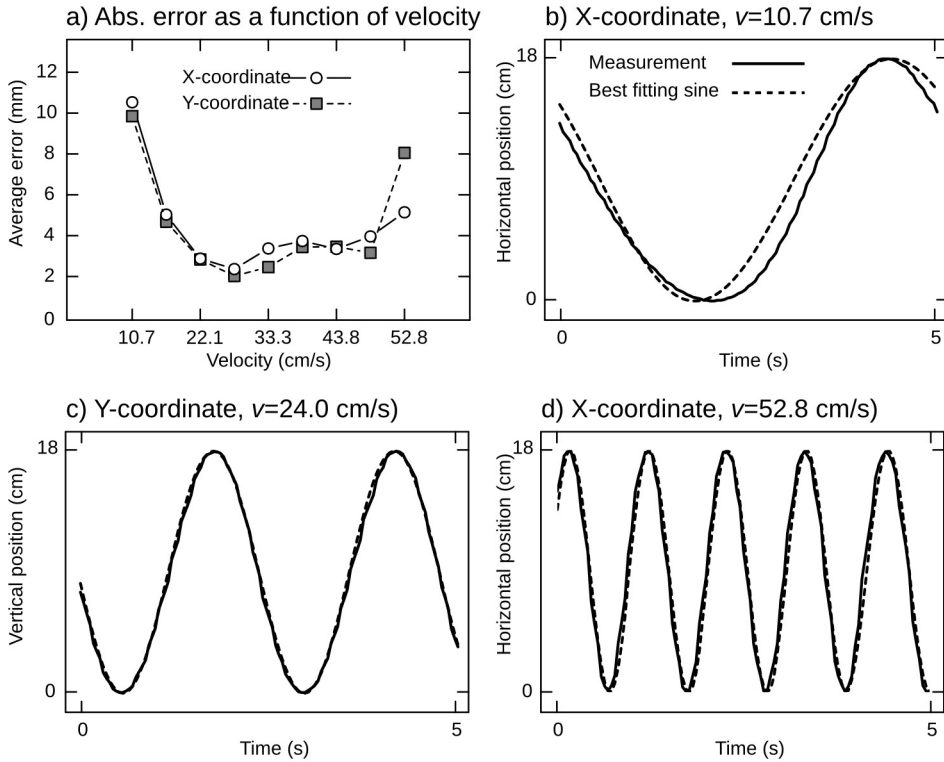
Figure x.7. Results of Experiment 3. a) The absolute measurement error of Mantra as a function of stimulus velocity. The spatial precision of Mantra varies from about 0.3° (2 mm) to 1.4° (10 mm). b) The first 5 seconds of unsmoothed X-coordinate measurements in the 10.7 cm/s speed level (blue line). The orange line is the best fitting sine. c) Same as (b), but for the Y-coordinate in the 24.0 cm/s speed level. d) Same as (b), but for the X-coordinate in the 52.8 cm/s speed level.

Results and discussion

The measurement error varied from about 0.3° (2 mm) to 1.4° (10 mm; Figure x.7a). Precision was highest for intermediate stimulus velocities (22.1 cm/s to 48.6 cm/s).

The results clearly show that Mantra's spatial precision is high, up to 0.3° for intermediate velocities. The pronounced precision dip for low stimulus velocities was unexpected. However, the reason for this anomaly becomes apparent when we inspect the unsmoothed data (Figure x.7b). For low stimulus velocities, the weight of the stimulus slows the movement down when the stimulus is on the way up (i.e., the lower part of the curve is wider than the upper part of the curve), causing a small deviation from a perfectly circular movement, and thus confounding our measure of precision. This issue does not (or very little) affect higher stimulus velocities (Figure x.7c,d), which therefore provide a

better estimate of Mantra's precision. We conclude that Mantra is able to track stimuli with a spatial precision of up to 0.3°. This corresponds to a precision of up to 2 mm if the camera is positioned at 40 cm from the stimulus, which is a typical distance.

Experiment 4

In Experiment 3 we could not directly compare the measured position of the tracked stimulus to its true position. Instead, we relied on an indirect measure: The assumption that the trajectory of the stimulus was, to a good approximation, circular.

Therefore, the aim of Experiment 4 was to extend the results of Experiment 3, using a paradigm in which the true position of the tracked stimuli is known. To this end, we tracked two stimuli that were presented on a computer display. The distance between the stimuli was varied and, after calibrating Mantra on a single stimulus configuration, we quantified the accuracy with which Mantra was able to measure the distance between the two stimuli.

Method

Stimulus, apparatus and procedure

The experiment was run on a laptop running Linux (Intel Core i3-i370m, 2.4 Ghz, Ubuntu 10.10) and created in OpenSesame 0.22 (Mathôt, Schreij, et al., 2012), using the Mantra Python bindings. Stimuli were presented on an external 19" TFT monitor, with a resolution of 1440x900px. Mantra 0.3 was run on the same laptop. For image acquisition a Trust Spotlight webcam was used, with a frame rate of 25Hz and a resolution of 640x480px. The webcam was placed in front of the external monitor, at a distance of 50cm.

Two circles ($r = 0.34°$), one purple and one green, were presented against a grey background, and were tracked using Mantra. The distance between the stimuli was varied in 28 steps from 0.34° to 9.8°. In addition, the stimulus configuration was rotated around the centre of the display in steps of 30° (angular). This way, stimuli were presented across the full field of view of the webcam, and distortions in the monitor as well as the webcam (such as pixels not being perfectly square), which are likely to affect measurements in a realistic experimental setting, were taken into account. 500 samples were recorded for each rotational step, yielding a total of 6000 samples for each distance.

Data analysis

We assumed that the real distance between the two stimuli was the raw distance measured by Mantra multiplied by a constant scaling factor. We calibrated Mantra (i.e.,
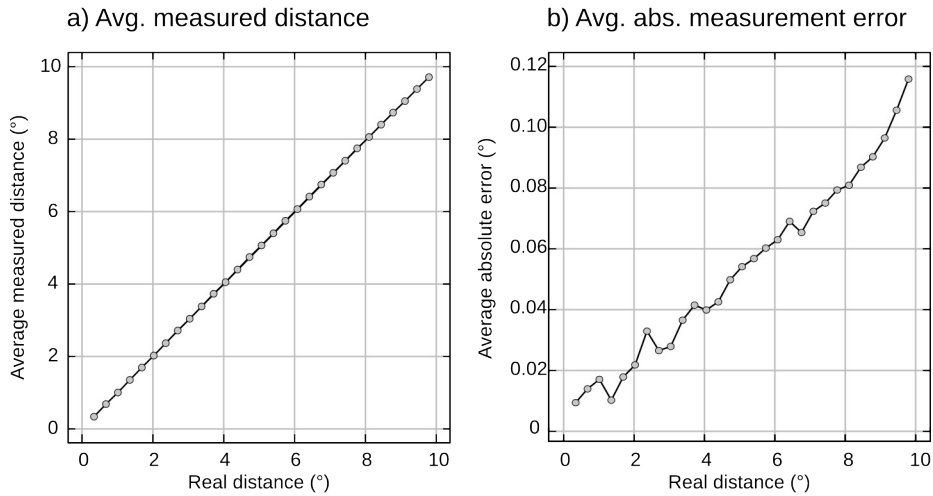
a) Avg. measured distance

b) Avg. abs. measurement error

Figure x.8. Results of Experiment 4. a) The distance between two stimuli as measured by Mantra, as a function of the real distance. b) The average absolute measurement error, as a function of the real distance. A more detailed overview of this data is presented in Table x.1.

determined the scaling factor) on an intermediate distance (5.07°). Next, we compared how well the calibrated distance measured by Mantra matched the real distance between the two stimuli. We determined the accuracy when averaged over all samples (i.e., the average error) as well as the accuracy for single samples (i.e., the average absolute error). In addition, we determined the standard error of the measurement. These analyses were performed separately for each distance level. No further statistics were performed, as the data is essentially descriptive.

Results and discussion

The results are shown in Figure x.8 and, in more detail, in Table x.1. Depending on the nature of an experiment it is important to know the measurement accuracy averaged over a large number of samples (for example when you are tracking a slow moving or static stimulus) or to know the measurement accuracy for single samples (for example when you are tracking a fast moving stimulus).

When looking at the average over 6000 samples, measurement error ranged from 0.00060° (for the 0.34° distance) to 0.082° (for the 9.8° distance). For single samples accuracy was also high, with an average absolute measurement error ranging from 0.0092° (for the 0.34° distance) to 0.12° (for the 9.8°) distance (Figure x.8b). This was also reflected by the standard error of the measurement, which ranged from 0.00015° (for the

0.33° distance) to 0.0014° (for the 9.8° distance). No anomalies occurred (such as tracking being lost or grossly inaccurate) during the entire session of 168,000 samples.

Table x.1. Results of Experiment 4. The distance between two stimuli as measured by Mantra, compared with the real distance.

Real distance (°)	Measured distance (°)	Standard error (°)	Average error (°)	Average absolute error (°)
0.33767	0.33707	0.00015	0.00060	0.00919
0.67533	0.68553	0.00017	0.01020	0.01372
1.01300	1.00590	0.00023	0.00710	0.01685
1.35067	1.35379	0.00020	0.00312	0.01001
1.68833	1.69635	0.00023	0.00801	0.01762
2.02600	2.02522	0.00033	0.00078	0.02161
2.36367	2.36686	0.00045	0.00320	0.03271
2.70133	2.71578	0.00035	0.01445	0.02633
3.03900	3.04210	0.00044	0.00310	0.02768
3.37667	3.38261	0.00050	0.00594	0.03632
3.71433	3.73203	0.00056	0.01769	0.04128
4.05200	4.05137	0.00055	0.00063	0.03967
4.38967	4.39983	0.00061	0.01016	0.04235
4.72733	4.74539	0.00067	0.01806	0.04960
* 5.06500	5.06500	0.00081	0.00000	0.05392
5.40267	5.40158	0.00080	0.00109	0.05656
5.74033	5.74603	0.00087	0.00569	0.06003
6.07800	6.06899	0.00088	0.00902	0.06277
6.41567	6.41357	0.00102	0.00210	0.06882
6.75333	6.74628	0.00094	0.00705	0.06515
7.09100	7.07187	0.00103	0.01913	0.07213
7.42867	7.4036	0.00107	0.02507	0.07483
7.76633	7.74628	0.00111	0.02005	0.07913
8.10400	8.06083	0.00111	0.04318	0.08071
8.44167	8.40065	0.00124	0.04102	0.08662
8.77933	8.73192	0.00132	0.04742	0.09005
9.11700	9.05095	0.00123	0.06605	0.09624
9.45467	9.38483	0.00135	0.06983	0.10537

* Point of calibration

General discussion

In the present paper we have introduced Mantra, a system for object-tracking. Mantra differs from existing object-tracking systems in three important respects. First, Mantra is freely available under an open-source license. Second, Mantra does not require expensive dedicated hardware. A computer and a camera (we used ordinary webcams for our experiments) are all that is needed to run Mantra. Third, Mantra is designed specifically as a tool for experimental psychology. Therefore, Mantra integrates well with E-Prime and Python. Mantra can be used from within other programming languages as well, provided that they have basic networking capabilities. This requires some additional coding for which the E-Prime and Python libraries can be used as templates.

In Experiment 1 we validated the spatial precision of Mantra in a realistic experimental setting, by replicating the Müller-Lyer illusion (Müller-Lyer, 1889). In this experiment participants matched two line-segments surrounded by inwards or outwards pointing arrowheads. Participants manipulated the length of one of the line-segments by changing the distance between their thumb and index-finger, which were tracked using Mantra. In Experiment 2 we validated the temporal precision of Mantra by using a variant of the additional singleton paradigm (Theeuwes, 1994). In this experiment participants reported the orientation of a target line-segment. In one condition participants responded using a keyboard. In another condition they responded by moving their index-finger, which was tracked using Mantra. Crucially, both methods of response yielded very similar results and there was no evidence for an increased level of noise when responses were collected using Mantra. In Experiments 3 and 4 we investigated the spatial precision and accuracy of Mantra more rigorously by tracking respectively a computer controlled physical stimulus and stimuli presented on a computer display. These experiments showed that under optimal conditions (i.e., tracking an artificial stimulus on a computer display) it is possible to track stimuli with a measurement error of less than 0.1°. Perhaps more realistically, under good conditions (i.e., tracking a properly defined physical stimulus) it is feasible to track stimuli with a spatial precision of up to 0.3°, which corresponds to about 2mm in a typical experimental set-up.

In summary, Mantra is a basic, but reliable and accurate object-tracking system. Mantra is freely available and has been designed specifically for use in psychological experiments. Because Mantra requires only a computer and a camera, it is possible to create a highly mobile experimental set-up. Mantra is unique in that it makes object-tracking accessible and easy to use for everyone.

∞

XI. A Simple Way to Estimate Similarity Between Pairs of Eye Movement Sequences

Abstract — We propose a novel algorithm to estimate the similarity between a pair of eye movement sequences. The proposed algorithm relies on a straight-forward geometric representation of eye movement data. The algorithm is considerably simpler to implement and apply than existing similarity measures, and is particularly suited for exploratory analyses. To validate the algorithm, we conducted a benchmark experiment using realistic artificial eye movement data. Based on similarity ratings obtained from the proposed algorithm, we defined two clusters in an unlabelled set of eye movement sequences. As a measure of the algorithm's sensitivity, we quantified the extent to which these data-driven clusters matched two pre-defined groups (i.e., the 'real' clusters). The same analysis was performed using two other, commonly used similarity measures. The results show that the proposed algorithm is a viable similarity measure.

∞

Now that affordable eye trackers have become commonplace (e.g., San Agustin et al., 2010), measuring eye movements is straight-forward. Given the right question, eye movements can provide insight into the inner workings of the mind. No wonder, therefore, that eye tracking is a popular tool among neuroscientists and psychologists.

But the apparent simplicity of conducting eye movement research is deceiving. Collecting large amounts of data is easy, but analysing the data in a way that does justice to the wealth of information they contain is emphatically not. Therefore, it is paramount that new, powerful tools for the analysis of eye movement data are developed. In the present paper, we will focus on one type of analysis in particular: the similarity measure. In the sense intended here, a similarity measure takes two eye movement sequences as input and returns a value, or set of values, that reflect how similar these sequences are. Such

similarity ratings can, in turn, be used as a starting point for more complex analyses. For example, in combination with clustering techniques, a similarity measure can be used to cluster eye movement sequences into more-or-less homogeneous groups.

In the sections that follow, we will highlight the importance of similarity measures for eye movement research, and review currently available methods. Next, we will propose a novel method, which we have called 'Eyenalysis'. We will argue that, particularly for exploratory analyses, this method has significant advantages over currently available methods and is considerably less complex. Finally, in an experiment using artificial, yet realistic eye movement data, we will show that the proposed method can be more sensitive than two commonly used alternative methods.

What a similarity measure can and cannot do for you

A similarity measure provides a way to answer a specific, but very common class of research questions. Broadly speaking, it allows you to cluster similar eye movement sequences together, or detect differences between predefined groups of eye movement sequences. We will illustrate this with a number of hypothetical research questions.

Question 1 (detecting differences between predefined sets)

"I have two sets of eye movement sequences, collected in two different experimental conditions. Are the two sets different from each other, in which case my experimental manipulation was effective?" Questions of this type can be answered by investigating whether the average similarity of pairs of eye movement sequences within sets is larger than the similarity of sequence pairs between sets.

Question 2 (diagnostic use)

"I have two sets of eye movement sequences. If I collect a single new eye movement sequence, can I determine to which of the two sets it belongs?" This question can be answered by determining which of the two sets is, on average (i.e., averaged over individual eye movement sequences) most similar to the 'target' sequence.

Question 3 (data-driven clustering)

"I have a large set of (unlabelled) eye movement sequences. I suspect that there are two distinct clusters hidden in this set. Can I detect these clusters in a data-driven way?" This question is similar to Question 1, but more stringent, because it does not require any a priori group-segmentation. The trick to solving this problem is to cross-compare all eye movement sequences and perform a cluster analysis on the resulting similarity matrix (cf.

Cristino et al., 2010). This type of analysis requires a highly sensitive similarity measure, and is the approach that we will use for the benchmark experiment described in the present paper.

Question 4 (within- versus between-subject similarity)

"How can I tell whether my data supports scanpath theory (Noton & Stark, 1971)? That is, are eye movement sequences of a person relatively constant across multiple viewings of the same scene?" This question can be answered by determining whether two sequences of the same person viewing the same scene are, on average, more similar than two sequences of the same person viewing different scenes and two sequences of different people viewing the same scene.

Similarity measures also have an important limitation: It is difficult to determine *why* two eye movement sequences show a particular degree of similarity. For example, using a similarity measure you may find that participants in Group A differ, with respect to their eye movements, from those in Group B. But it is difficult to specify in which regard these two groups differ. This does not hold equally strongly for all approaches. For example, when using the Levenshtein distance (1966; see the section on *Existing similarity measures*) you can inspect the relative frequency of omissions and substitutions. Similarly, the approach by Jarodzka, Holmqvist, and Nyström (2010) allows you to compare similarity ratings across a number of dimensions to get some insight into the 'why' question. But in all cases this insight is limited. This is important to bear in mind when considering a similarity measure for use as part of an analysis.

Existing similarity measures

Similarity measures have a venerable tradition in eye movement research, and many variations on this common theme have been tried. Broadly speaking, there are three dominant approaches: similarity measures based on correlations between 'attention maps' (Caldara & Miellet, 2011; Gibboni, Zimmerman, & Gothard, 2009; Grindinger et al., 2011), string edit methods (Brandt & Stark, 1997; Cristino et al., 2010; Foulsham & Underwood, 2008; Hacisalihzade et al., 1992; Levenshtein, 1966; West, Haake, Rozanski, & Karn, 2006; Zangemeister & Oechsner, 1996), and various geometric methods (Dempere-Marco, Hu, Ellis, Hansell, & Yang, 2006; Henderson, Brockmole, Castelhano, & Mack, 2007; Jarodzka et al., 2010; Mannan, Ruddock, & Wooding, 1995, 1997; Zangemeister & Oechsner, 1996).
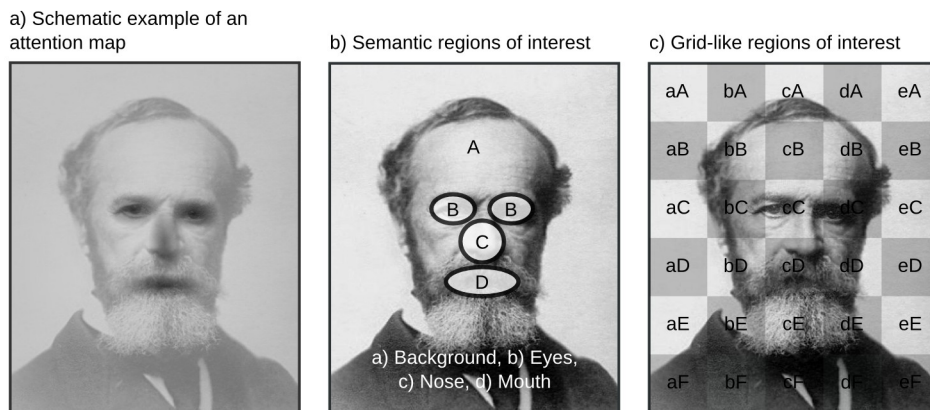
a) Schematic example of an attention map   b) Semantic regions of interest   c) Grid-like regions of interest

Figure xi.1. Alternative ways to determine the similarity between eye movement sequences. a) Fixation density can be plotted as an attention map (Caldara & Miellet, 2011). Correlations between attention maps can be used as a measure of similarity. b,c) An image can be divided into regions of interest (RoIs) based on the semantic properties of the image (b) or based on a grid (c). Using these RoIs, eye movement sequences can be re-coded as character strings, and a string edit distance can be used as a similarity measure (Cristino et al., 2010; Levenshtein, 1966; West et al., 2006).

Attention maps

Attention (or fixation) maps are heat maps in which hotspots correspond to frequently fixated areas, or areas with a high total fixation duration (for a sophisticated implementation, see Caldara & Miellet, 2011). For example, an attention map based on the eye movements of participants viewing pictures of faces will typically contain hotspots surrounding the eyes, nose, and mouth (Figure xi.1a). Although there are different ways to derive similarity from attention maps, the general idea is straight-forward: If two attention maps are strongly correlated, they reflect highly similar eye movement sequences.

The downside of attention maps is that they contain no representation of fixation order. One can circumvent this limitation by analysing subsequent time-windows separately (e.g., Grindinger et al., 2011). But, from a practical point of view, the minimum size of the time-window is constrained by the need to maintain a sufficient number of fixations in each temporal bin. Therefore, attention maps are, in most cases, sub-optimal if one is interested in the temporal properties of eye movement sequences.

String edit methods

String edit methods are traditionally the most common way to determine the similarity between eye movement sequences (Brandt & Stark, 1997; Cristino et al., 2010; Duchowski et al., 2010; Foulsham & Underwood, 2008; Hacisalihzade et al., 1992; Privitera & Stark,

2000; West et al., 2006; Zangemeister & Oechsner, 1996). In this approach, pioneered by Hacisalihzade, Stark, and Allen (1992), eye movement sequences are re-coded as character strings. In order to make this possible, the image is segregated into different regions of interest (RoIs). This can be done based on the semantic properties of the image (Figure xi.1b). For example, for the picture of a face it would make sense to divide the image into at least four RoIs, corresponding to the eyes, nose, mouth, and background respectively. Alternatively, the image can be divided into a grid, in which case no assumptions have to be made about the most sensible semantic segregation of the image (Figure xi.1c). Finally, some authors have proposed a data-driven way to define RoIs automatically. This can be done post-hoc, based on the viewing patterns of the participants, or beforehand, based on an analysis of the image (e.g., Privitera & Stark, 2000).

The next step is to re-code the eye movement sequence as a string of characters. Let's consider the following eye movement sequence:

```
eyes → nose → mouth → eyes
```

Given the RoIs from Figure xi.1b, the corresponding character string would be:

```
BCDB
```

After re-coding, all that is needed is a suitable similarity measure for character strings, for which there are many well-established algorithms. The best known of these are the Levenshtein distance (Levenshtein, 1966) and its numerous variations (Okuda, Tanaka, & Kasai, 1976; Wagner & Lowrance, 1975; Zangemeister & Liman, 2007).

In its simplest form, (i.e., the unmodified Levenshtein distance; Levenshtein, 1966), the string edit method suffers from a number of severe drawbacks. Specifically, it does not take into account factors such as fixation duration, nor the fact that RoIs are usually not 'equally unequal' (e.g., given the RoIs from Figure xi.1c, 'aA' is more similar to 'bA' than to 'cF'). We have recently proposed a string edit method, based on the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970), in which most of these issues have been resolved (Cristino et al., 2010). This new method, which we called 'ScanMatch', is substantially more sensitive than the traditional string edit methods. But there are more general concerns that are not easily addressed within the constraints imposed by the string edit framework.

For example, any string edit method requires an image to be divided into RoIs. It can be difficult, or prohibitively time consuming, to define semantic RoIs (Figure xi.2), and the validity of data-driven/ artificial RoIs (Privitera & Stark, 2000) has been questioned (Grindinger et al., 2011). As a result, some researchers prefer to use grid-like RoIs (Figure xi.1c). In this case the RoIs serve as a proxy for a low-resolution coordinate system, and there may be significant advantages to using a more natural, geometric representation.

Figure xi.2. The LabelMe website allows visitors to define semantic RoIs (indicated by the coloured areas in the right pane) in an image (Russel, Torralba, & Murphy, 2008), thus using crowd-sourcing to overcome the difficulties inherent to semantic RoIs.

Geometric methods

In geometric (or dimensional) methods, eye movements are represented by their geometric properties (location, saccade direction, fixation duration, etc.). This stands in contrast with the statistical approach of attention maps, and the RoI approach of string edit methods.

Zangemeister and Oechsner (1996) and, more recently, Jarodzka et al. (2010) have proposed algorithms that are essentially intermediates between string edit methods and geometric methods. In this approach, eye movement sequences are represented by series of vectors that represent (usually) the direction and amplitude of a saccade. But the approach is similar in spirit to string editing through its use of alignment (cf. Needleman & Wunsch, 1970): Series of vectors that line up well are considered similar. This method has the advantage of doing away with the awkward need for RoIs and re-coding schemes. In addition, Jarodzka et al.'s (2010) algorithm has an interesting property: It allows researchers to determine different similarity measures that each focus on a different aspect of the eye movements (e.g., shape, position, or length). Whether this is a feature or a limitation depends on the goals and prior knowledge of the researcher. It is a feature when a researcher has a specific hypothesis about the dimensions that he or she expects to be most relevant. It is a limitation in exploratory research, when a firm hypothesis is lacking.

Mannan et al. (1995, 1997; see also Henderson et al., 2007) have proposed a 'nearest neighbour' method that is conceptually most similar to the method that we will propose in the present paper, albeit less flexible. Mannan et al. (1995) represent eye movement sequences as sets of fixations (i.e., x, y coordinate pairs). Each fixation is mapped onto the nearest fixation from the other set. This results in a large number of mappings, each

171

associated with a mapping distance. The (overall) distance is the sum of all mapping distances (after normalising for the length of the eye movement sequences).

A clever variation on this approach has been described by Dempere-Marco et al. (2006), who used the earth mover distance (EMD) or Wasserstein metric. The EMD is generally conceptualized as the amount of traffic that is required to fill a set of holes (the fixations in sequence A) with a set of dirt piles (the fixations in sequence B). The advantage of this approach over a point-mapping rule, such as the one used by Mannan et al. (1995), is that it allows one to take fixation duration into account: Long fixations correspond to deep holes or large piles of dirt.

The methods of Mannan et al. (1995) and Dempere-Marco et al. (2006) do not require re-coding and RoIs. However, the drawback of these methods is that they do not take fixation order into account. The similarity measure that we will propose here can be viewed as a simplified, multidimensional variation on the method developed by Mannan et al. (1995, 1997).

The proposed similarity measure

Representation

Sets of fixations

We represent eye movements sequences as sets of fixations. Each fixation is defined by an arbitrary number of dimensions. For example, a fixation may be defined only by its location, in which case it has two dimensions (x, y). (Assuming that we do not take vergence into account, otherwise there would be a z dimension as well.) But in principle any number and combination of dimensions can be used, which is the primary departure from Mannan et al.'s method (1995, 1997). For example, in many situations it would make sense to define fixations by their location, timestamp and duration, in which case there would be four dimensions (x, y, t, d). Note that, unlike in Jarodzaka et al.'s (2010) method, the set of fixations is unordered. Nevertheless, the temporal properties of an eye movement sequence can be readily taken into account by incorporating temporal dimensions such as time and fixation duration.

Using eye tracker output

The benefit of this representation is that it closely matches the output from most eye trackers, which generally (although not always) offer an abstraction layer in which individual gaze samples are converted into larger-scale events, such as fixations and

saccades. For example, in the raw data produced by the Eyelink series of eye trackers (SR Research, Mississauga, ON, Canada), fixations look like this:

```
EFIX L 16891857 16893183 1327 32.7 369.2 8588
```

Or, more schematically:

```
EFIX [eye] [start time] [end time] [duration] [x] [y] [pupil size]
```

The relevant dimensions can be easily extracted from this type of raw data, and no elaborate re-coding scheme will usually be required.

Data 'whitening'

However, one situation in which some pre-processing is required is when you want to incorporate dimensions that lie on qualitatively different scales.

To illustrate this point, let's consider the following example: We use location and fixation duration (x, y, d) as dimensions. We use seconds as units for d and pixels as units for x and y. This means that values for d will generally be small (below one), whereas values for x and y will be large (range in the hundreds). More precisely, the problem is that d has less variance than x and y. Because of this imbalance, d will contribute little to the distance measure.

This problem can be resolved through a process called 'whitening': For each dimension, all values are divided by the standard deviation of values within that dimension. As a result of this scaling operation, all dimensions will have unit variance, and will contribute equally to the distance measure.

It is difficult to say whether or not whitening should be applied in a given situation, because it is not necessarily beneficial when applied inappropriately. It may be desirable for some dimensions to have a relative large variance. For example, when you increase the length of an eye movement sequence, the variance in time (t) will increase, but the variance in position (x, y) may not. In this case, the difference in variance between dimensions may be informative, and should not be undone through whitening. Conversely, the value on a particular dimension may be essentially constant (for example the y coordinate if participants are following a horizontally moving dot), except for noise. If this is the case, whitening is undesirable, because it will have the effect of amplifying noise.

Given these considerations, we propose, as a rule of thumb, not to apply whitening unless some dimensions are obviously incomparable (i.e., the standard deviation differs more than an order of a magnitude between dimensions), or if there is a theoretical reason why variance should be strictly equal across dimensions.

Distance measure

    Rationale

The goal of the proposed distance measure is to take two eye movement sequences, which we will call *S* and *T*, and return a value that estimates the distance (i.e., the inverse of the similarity) between *S* and *T*.

We propose that the best way to achieve this is by constructing a mapping between *S* and *T*, so that each point (i.e., fixation) from *S* is mapped onto at least one point from *T*, and vice versa. The goal is to minimize the (normalised) sum of the distances associated with all mappings (Figure xi.3a).

This 'mapping problem' has no known solution that is both efficient and guaranteed to be optimal, but there are various heuristic that consistently achieve a very good mapping. In preliminary analyses we have explored a number of different heuristics and have found that 'double-mapping' is the preferred technique, because it is computationally cheap and not notably, if at all, less accurate than more sophisticated heuristics (cf. Mannan et al., 1995).

In the double mapping technique, each point from *S* is mapped onto the nearest neighbour from *T*. In addition, each point from *T* is mapped onto the nearest neighbour from *S* (Figure xi.3b). Many mappings thus occur twice. Importantly, double mapping does not suffer from complex problems such as the need to split long mappings into multiple shorter ones, or pruning of spurious mappings.

As noted by Henderson et al. (2007), double-mapping has the risk of mapping a large number of points from *S* onto a single point (or small number of points) from *T*. This is true, but in general we prefer the double-mapping approach over the 'unique assessment' mapping rule proposed by Henderson et al. (2007). This is because, unlike unique assessment, double-mapping does not require an equal number of points in each set (i.e., eye movement sequences of different lengths can be compared), and therefore allows for a broader application.

    Algorithm

A point-mapping is the mapping between a point *p* in *S* and a point *q* in *T*, and is associated with a distance, *d(p,q)*, which is the Euclidean distance between *p* and *q*:
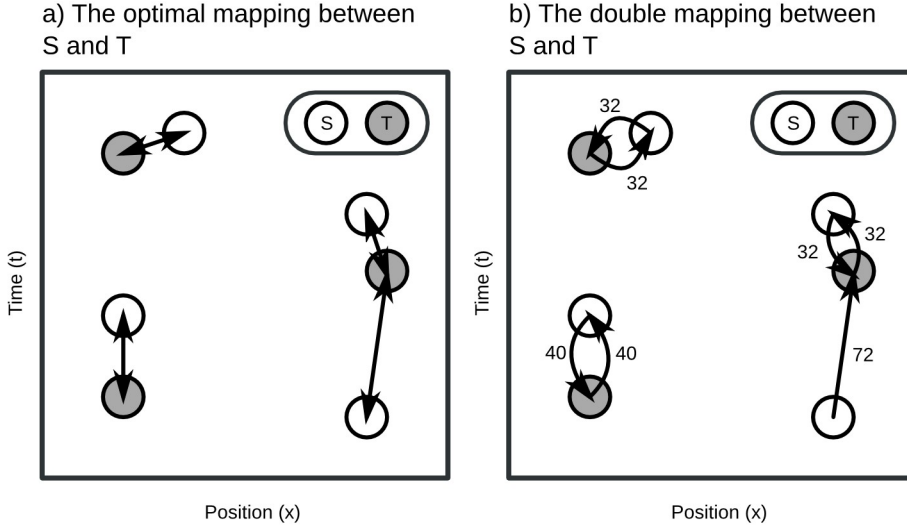
Figure xi.3. A schematic illustration of the mapping principle. For display purposes, only two dimensions (x, t) are shown, but the principle generalizes to an arbitrary number of dimensions (Formula xi.1). a) The optimal mapping between S and T. b) The double mapping, which is a good and computationally cheap approximation of the optimal mapping. In this example, we can determine the distance between S and T as follows (applying Formula xi.2): D(S,T)=(32+32+40+40+32+32+72)/max(3,4), D(S,T)=70

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2}$$

Formula xi.6

Here, n is the number of dimensions, and p_i and q_i are the i-th dimension of p and q, respectively.

A sequence-mapping between *S* and *T* is the collection of all point-mappings. Following the double-mapping technique, all points from *S* are mapped onto their nearest neighbour in *T*, and vice versa. A sequence-mapping is also associated with a distance, *D(S,T)*, which is the normalised sum of all the point-mapping distances. Normalisation occurs by dividing *D(S,T)* by the number of points in the largest sequence. This prevents long sequences from being unfairly penalised:

$$D(S,T) = \frac{\sum_{i=1}^{n_S} d_S^i + \sum_{j=1}^{n_T} d_T^j}{max(n_S, n_T)}$$

Formula xi.7

Here, n_S is the length of S, n_T is the length of T, d_S^i is the distance between point *i* in S to its nearest neighbour in T, d_T^j is the distance between point *j* in T to its nearest neighbour in S, and D(S,T) is the distance between S and T.

A different, and perhaps more intuitive, way of describing the algorithm is through pseudo-code. The equivalent pseudo-code is as follows:

```
D = 0
For all points p in S:
      Find nearest point q in T
      D = D + distance(p,q)
For all points q in T:
      Find nearest point p in S
      D = D + distance(p,q)
D = D / max(size(S),size(T))
```

Again, *S* and *T* denote two eye movement sequences, *p* and *q* denote points in *S* and *T* respectively, *distance()* is the Euclidean distance function, and *D* is the resulting distance.

Implementation

We have developed an optimized Python (Jones et al., 2001; Van Rossum, 2008) implementation of the algorithm, which can be obtained from the first author's website. In addition to the algorithm per se, this implementation provides functionality for reading text data, whitening data, cross-comparing large datasets, and performing *k*-means cluster analyses. Documentation and demonstration scripts are included.

As of yet, the algorithm has not been implemented in other programming languages. However, as is apparent from the pseudo-code shown above, implementing the algorithm is trivial in most languages, particularly those that have strong matrix- and data-manipulation capabilities, such as Python, R (R Development Core Team, 2010), and MATLAB (The MathWorks, 1998) / Octave (Eaton, 2002).

Effects of sequence length, dimensionality, and spacing

An important limitation to keep in mind when applying a distance measure, such as the one proposed here, is that distance ratings are only meaningful within a particular set of data—Distance ratings do not have an absolute meaning.

To illustrate this, we calculated the mean distance between randomly generated sequences (N=1000 for each data-point). This was done with various numbers of dimensions (1, 2, 4, 8, and 16) and with various sequence lengths (1-32). We also varied 'fixation spacing', by which we mean the following: In the *fixed space* condition (Figure xi.4a) coordinates had random values between 0 and 1. In the *growing space* condition (Figure xi.4b), coordinates had random values between 0 and *X,* where *X* was equal to the sequence length. In other words, the *fixed space* condition simulated a situation in which gaze is strictly confined, whereas the *growing space* condition simulated a situation in which the eyes roam free, inspecting an ever growing area.

The effect of dimensionality is clear (Figure xi.4). Increasing the number of dimensions leads to higher distance ratings. This is not surprising, because, in a sense, the Euclidean distance function (Formula xi.1) does not fully normalise for dimensionality. This can be intuitively shown with an example: Opposite corners of a cube (three dimensions) are further apart than opposite corners of a square (two dimensions), provided that the length of the edges is kept constant.

More surprising perhaps, is that the effect of sequence length is variable. Specifically, it depends on whether fixations are spaced within a fixed region (Figure xi.4a) or a region that expands as the number of fixations increases (Figure xi.4b). This is a result of the normalisation procedure (Formula xi.2). If fixations are spaced within a fixed region, normalisation over-compensates, and the mean distance rating decreases with increasing sequence length. If, on the other hand, fixations are spaced within in a region that grows as the number of fixations increases (Figure xi.4b), normalisation under-compensates and the mean distance rating increases with sequence length.

This simulation illustrates that normalisation for sequence length is inherently problematic. If the point of gaze is strictly confined within a fixed region, the optimal normalisation procedure is different from when gaze is allowed to roam completely free. In practice, one may observe any intermediate between these two extremes: As people scan an image, their eyes will sequentially inspect different locations, and thus the region that contains fixations will grow over time. But, at the same time, gaze is restricted by factors such as screen boundaries, so the region that contains fixations cannot grow indefinitely.
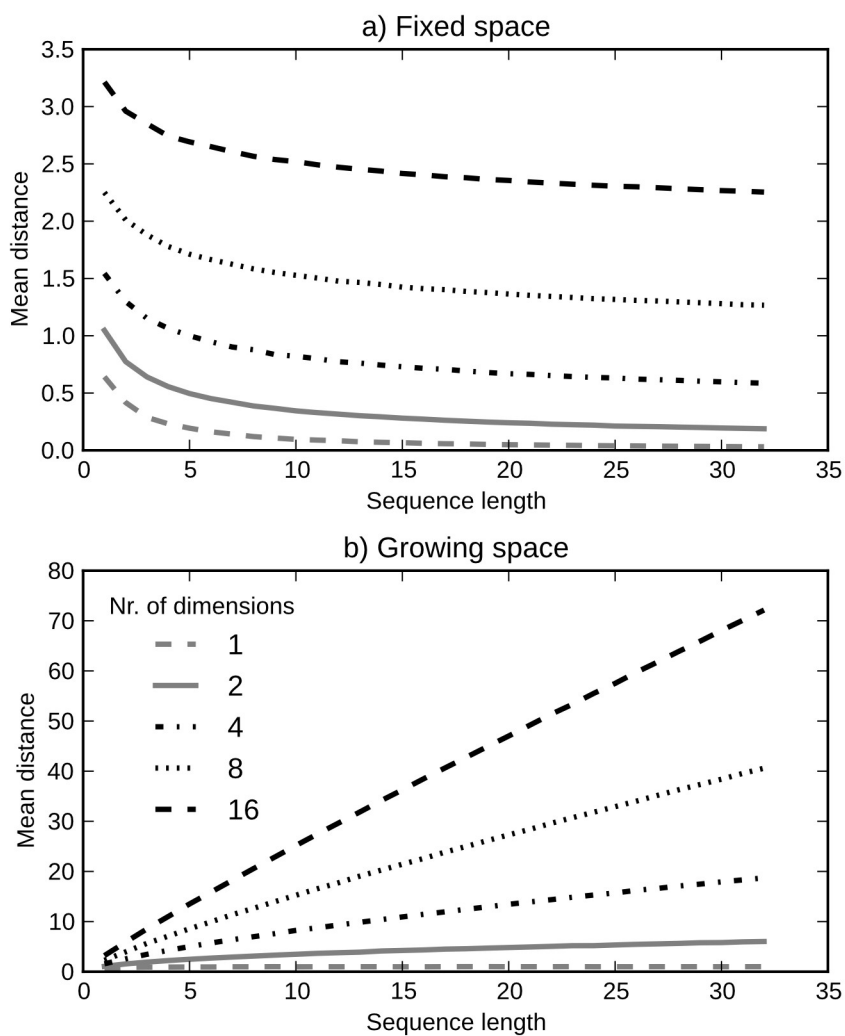
Figure xi.4. The mean distance between two randomly generated sequences as a function of dimensionality, sequence length, and spacing. a) Values were randomly chosen between 0 and 1. b) Values were randomly chosen between 0 and X, were X is equal to the sequence length.

In summary, distance ratings are relative and do not carry meaning outside of a particular dataset. With respect to the distance measure proposed here, mean distance ratings are affected by the number of dimensions, the (average) sequence length, and, more subtly, the way in which fixations are spread out over space, time, and other dimensions.

Experiment

We have conducted an experiment to compare the sensitivity of the proposed algorithm, Eyenalysis, to that of existing algorithms. Specifically, we compared Eyenalysis to ScanMatch (Cristino et al., 2010) and the Levenshtein distance (Levenshtein, 1966). The reason for choosing these two algorithms as points of reference is that they represent both the traditional (the Levenshtein distance) and state-of-the-art (ScanMatch) in similarity measures.

The term 'sensitivity' requires some clarification in this context. Essentially, we define sensitivity operationally as how well a similarity measure deals with noise in experiments such as the present one.

We generated a large number of artificial, yet realistic eye movement sequences that fell into two categories. Next, we performed a cross-comparison of this dataset (using a similarity measure), and performed a k-means cluster analysis on the resulting cross-comparison matrix. This yielded two clusters of eye movement sequences. Our measure of interest is how well the two clusters, which have been generated in a data-driven way, match the two given categories (i.e., the 'real' clustering).

In situations with very little noise (i.e., highly distinct categories) we expect any sensible similarity measure to perform perfectly. In situations with very high levels of noise, we expect any similarity measure to perform at chance level. However, the amount of noise that a similarity measure is able to cope with is taken as a measure of its sensitivity.

In the present experiment, the data-set is three-dimensional, containing the position (x, y) and time-stamp (t) of each fixation. We chose this representation, because it is a natural and common way to represent eye movement data, and because it allows for a straight-forward comparison to ScanMatch and the Levenshtein distance. However, in Eyenalysis all dimensions are treated in the same way, regardless of the type of information that they convey. So the labels that we have attached to the dimensions are, in a sense, arbitrary.

All scripts, input data, and output data are available from
http://www.cogsci.nl/eyenalysis.

Data generation procedure

Generating artificial eye movement sequences

As a starting point, we took two images (Figure xi.6a,b). Using the iLab Neuromorphic vision toolkit (Itti & Koch, 2000; Itti et al., 1998), we generated an artificial eye movement sequence, consisting of 10 saccades (11 fixations), for each of the two images. Each fixation was defined by a timestamp (*t*) and a position (*x*, *y*).

For 200 levels of noise (σ from 0 to 1990 in steps of 10; in px for *x*, *y*; in ms for *t*) we did the following: The two sequences were copied 50 times and noise was added to each copy (Figure xi.6c,d; Figure xi.5). Specifically, a random value (sampled from a normal distribution with $\mu = 0$ and σ per the noise level) was added to *x*, *y*, and *t* for all fixations. *x* was constrained between 0 and 1280 (the width of the images), y between 0 and 960 (the height of the images), and *t* between 0 and 5000. For each fixation there was a probability of σ/4000 of either an omission or an addition. An omission meant that the fixation was skipped. An addition meant that the fixation was followed by a new, completely random (within the given constraints) fixation.

Character string representation

Because the Levenshtein distance (Levenshtein, 1966) and ScanMatch (Cristino et al., 2010) require input in the form of character strings, the eye movement sequences were re-coded as character strings. Each fixation was coded as a pair of characters (e.g., aA), where the first character represents *x* and the second character represents *y*. This representation was chosen for compatibility with ScanMatch (Cristino et al., 2010). As described below, we used a slightly modified version of the Levenshtein distance (Levenshtein, 1966), to overcome its single-character (or 26 RoIs) limit. *t* was represented as repetition of a character-pair (Figure xi.5). For each 100ms, a character-pair was repeated. So, for example, a 350 millisecond fixation in the upper-left of the image would be represented as:

```
aAaAaA
```

Analysis

For each algorithm (ScanMatch, Levenshtein distance, and Eyenalysis) and noise level (0 to 1990) we performed the following analysis: Each movement sequence was compared to each other eye movement sequence. This resulted in a 100x100 matrix of distance scores. Using the PyCluster package (de Hoon, Imoto, & Miyano, 2010), a 1-pass *k*-means cluster analysis was performed on the cross-comparison matrix to obtain 2 clusters.
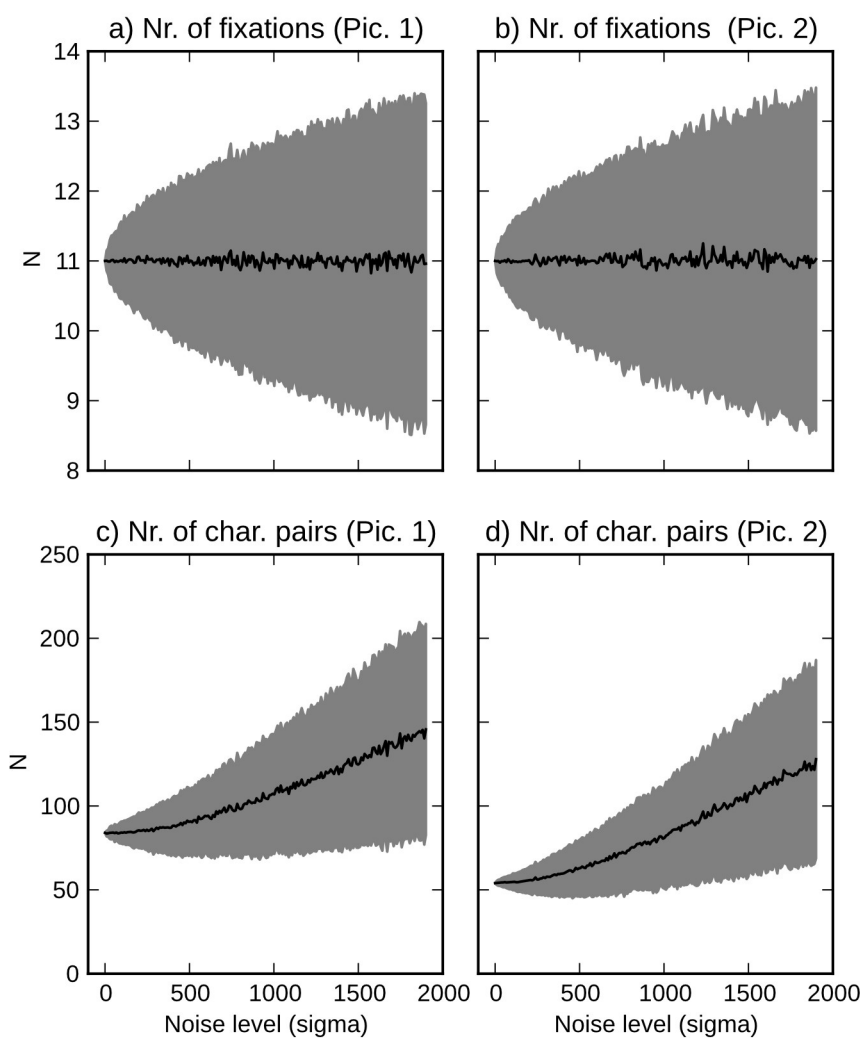
Figure xi.5. Descriptive statistics (the shaded area indicates the standard deviation) for the artificial eye movement data. Whereas the average number of fixations is relatively constant across noise levels and the two pictures, the average length of the character strings increases. This is because the length of the character strings also reflects the duration of the fixations.
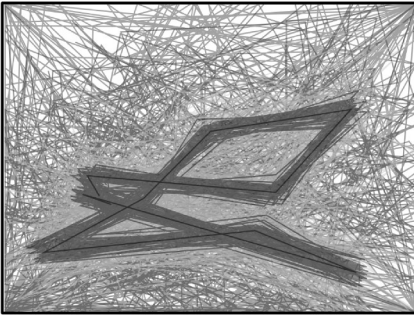
a) Image 1: *Eiffeltower and the Seine at night* (WikiMedia Commons)

b) Image 2: Theobald von Oer *Der Weimarer Musenhof* (1860)

c) Scanpaths with various levels of noise for Image 1

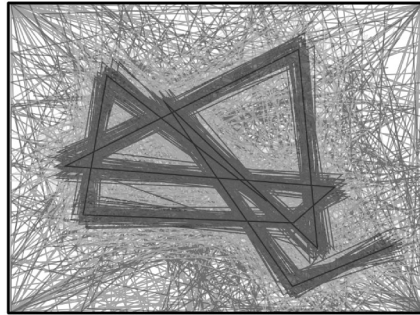d) Scanpaths with various levels of noise for Image 2

Figure xi.6. a,b) Two images were used to generate realistic, artificial eye movement data using the iLab Neuromorphic vision toolkit (Itti & Koch, 2000; Itti et al., 1998). c,d) Different levels of noise (indicated by different colours) were added to the eye movement sequences from (a,b).

Clustering accuracy and chance level

The clusters determined by *k*-means clustering are unlabelled, in the sense that it is not defined which cluster (*kmeans$_A$* or *kmeans$_B$*) matches which image (*image$_A$* or *image$_B$*). We therefore first determined the clustering accuracy assuming that *kmeans$_A$* maps onto *image$_A$*, and reversed this mapping if the clustering accuracy was less than 50%. Because this approach prevents accuracy from dropping below 50%, we needed to explicitly determine chance level. An analysis using random data set chance level at 54%.

Application of ScanMatch

A 26x26 grid ('number of bins') with an RoI modulus of 26 was used. A substitution matrix threshold of 19 was used, which was 2 times the standard deviation of the 'gridded'

saccade size (cf. Cristino et al., 2010). The gap value and temporal bin-size were left at 0. Because all parameters were either derived from the data in a predetermined manner, or left at their default value, there were no free parameters in our application of ScanMatch.

Application of Levenshtein distance

We used the classic Levenshtein distance (Levenshtein, 1966), with two modifications to allow for a more straight-forward comparison to the other algorithms. Firstly, we used character-pairs, rather than single characters, as units for matching. This was done so that we could use the same dataset as input for both ScanMatch (Cristino et al., 2010) and the Levenshtein distance. Secondly, the resulting distance-score was normalised by dividing the score by the length of the largest eye movement sequence. This normalisation procedure is not part of the Levenshtein distance per se, but is commonly applied when used in eye movement research (e.g. Foulsham & Underwood, 2008). There were no free parameters in our application of the Levenshtein distance.

Application of Eyenalysis

Eyenalysis was applied on both the raw dataset and on the whitened data, as outlined in the section *Data 'whitening'*. There were no free parameters in our application of Eyenalysis.

Results

The results of the experiment are shown in Figure xi.7. In line with Cristino et al. (2010), we found that the Levenshtein distance (Levenshtein, 1966), despite being a widely used method, offers poor performance relative to the other methods that we tested. In the current experiment, the simplest method was most sensitive: Eyenalyis applied on the raw dataset was more sensitive than both ScanMatch and Eyenalysis applied on the whitened data.

Discussion

We have proposed Eyenalysis, a novel algorithm to estimate the similarity between eye movement sequences. Using realistic artificial eye movement data, we have shown that Eyenalysis is more sensitive, at least in the present experiment, than the commonly used Levenshtein distance (Levenshtein, 1966) and ScanMatch (Cristino et al., 2010), an advanced string edit measure that we have previously proposed to overcome the limits of traditional string edit methods.

With an eye towards an application in real-life experimental settings, an important feature of Eyenalysis is its simplicity. Applying the algorithm is straight-forward, and does
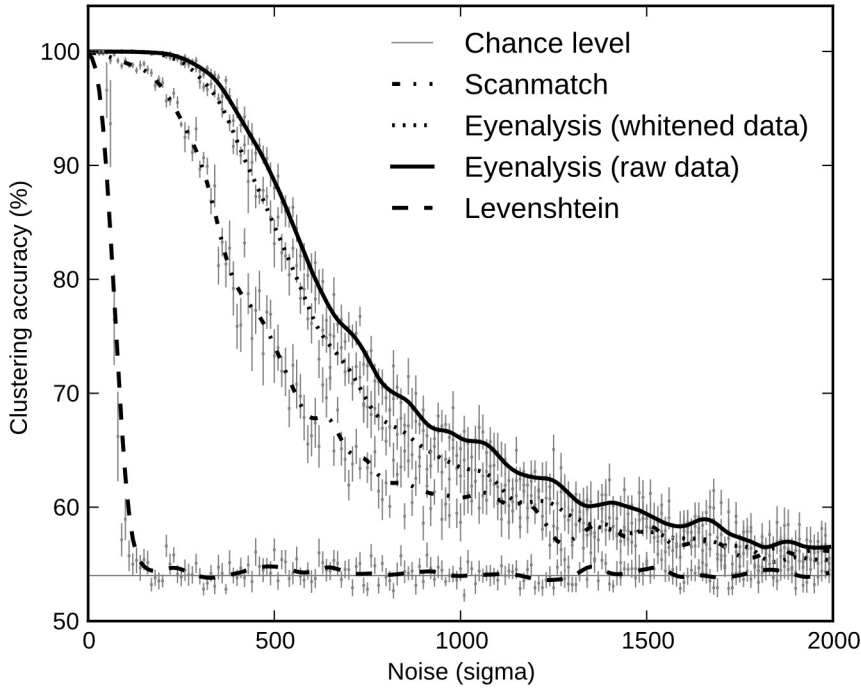
Figure xi.7. Results of the experiment. a) Clustering accuracy as a function of method and noise level. Error bars indicate the standard error. Lines have been smoothed using a Hanning window of width 70.

not require re-coding eye movement data into a special format, such as a character string representation. A Python implementation is provided, but the algorithm can easily be implemented from scratch in any programming language.

A landmark study by Noton and Stark (1971) illustrates how similarity measures can be used to elucidate theoretical issues. Noton and Stark (1971) noted that people tend to scan images in a stereotyped way. That is, the eye movement sequence of a person is relatively constant across multiple viewings of the same image (but not across different people, or across different images). Based on this finding, they proposed that eye movement sequences are an integral part of memory. By consistently viewing the same scene in (more or less) the same way, one can predict the visual input that is expected on each fixation. Therefore, so Noton and Stark (1971) argued, stereotyped eye movements could facilitate recognition.

Although their results were convincing, Noton and Stark (1971) did not perform a rigorous analysis. The similarities were obvious on visual inspection of the data (but see Privitera & Stark, 2000 for a more recent, quantitative corroboration). However, in some cases, for example when the data-set is large or noisy, a quantitative similarity measure, such as the one proposed here, is required. More specifically, a similarity measure can help researchers to address a particular, very common class of research questions. For example, one can estimate whether there are differences between predetermined groups of eye movement sequences (e.g., corresponding to different experimental conditions). Or, when combined with a cluster analysis, one can split a large set of eye movement data into groups of more-or-less homogeneous eye movement sequences in a data-driven way (also see Duchowski et al., 2010 and Privitera & Stark, 2000 for sophisticated similarity-based analyses).

The usefulness of similarity measures has been long recognised, and quite a few different methods have been proposed (Caldara & Miellet, 2011; Cristino et al., 2010; Dempere-Marco et al., 2006; Duchowski et al., 2010; Gibboni et al., 2009; Grindinger et al., 2011; Jarodzka et al., 2010; Levenshtein, 1966; Mannan et al., 1995, 1997; Privitera & Stark, 2000; West et al., 2006). Although some methods are more sensitive than others, many are useful in practice (Foulsham & Underwood, 2008; Henderson et al., 2007), and the choice for a specific algorithm depends largely on the goals of the researcher.

Eyenalyis is particularly well suited for exploratory analyses, because it allows one to simultaneously include many different factors in the analysis, and does not require the expected differences to be specified a priori. The algorithms proposed by Mannan et al. (1995, 1997; see also Henderson et al., 2007) and Dempere-Marco et al. (2006) are very similar to Eyenalysis when only positional information is considered. The primary contribution of Eyenalysis is to make it possible to include an arbitrary number and combination of dimensions. Any property of a fixation can be included in the analysis, provided that a numerical value can be assigned to it.
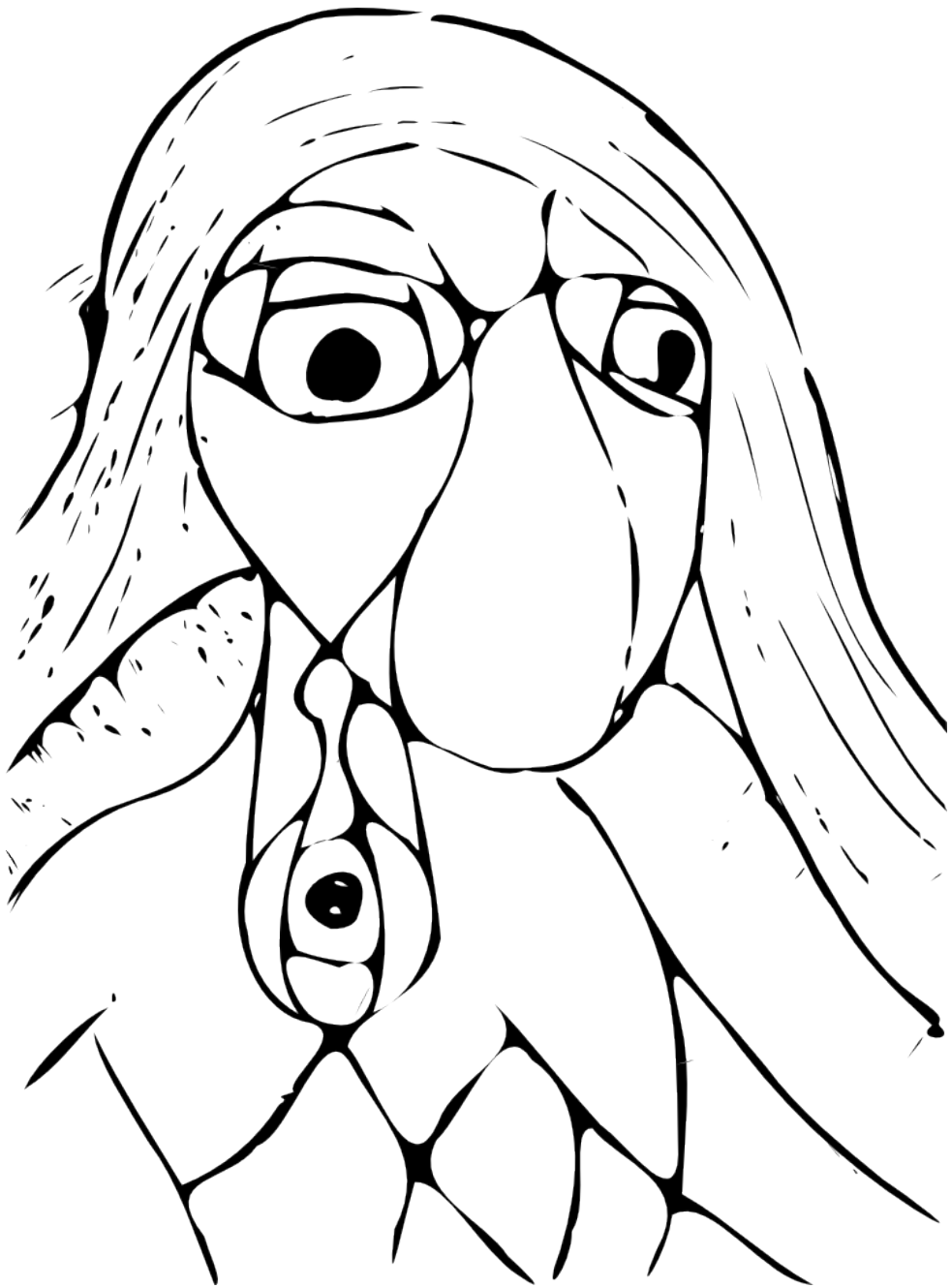
But there are also a number of limitations. As noted in the introduction, it is difficult to interpret similarity ratings obtained from Eyenalysis (and to some extent this is true of all similarity ratings). Consider, for example, an experiment in which you expect two groups to differ in the latencies of their saccadic eye movements. The problem with using a similarity measure in this case is that, even if you find a difference between the groups (i.e., eye movement sequences are more similar within groups than between groups), you cannot be sure that this difference is indeed driven primarily by a difference in saccadic latencies. Therefore, additional analyses may be required to interpret the similarity ratings.

Another limitation has to do with the relative weights that are assigned to each dimension (position, time, fixation duration, etc.). Weighting dimensions is straight-forward. If you want a dimension to exert a larger influence on the similarity rating, you multiply all values in that dimension by some factor larger than 1. Conversely, the importance of a dimension can be reduced by multiplying all values by a factor between 0 and 1. But the difficulty lies in deciding on an appropriate weighting. This is essentially a conceptual problem that revolves around the proper definition of 'similarity': Is a distance of 100px comparable to an interval of 10ms, 100ms, or 1000ms? At present, there is no satisfactory solution to the issue of dimensional weighting, particularly when dimensions with incomparable units (e.g., pixels and milliseconds) are incorporated. As a rule of thumb, we propose that the variance within dimensions should be kept relatively constant. If this is not the case, a 'whitening' procedure can be performed, as described in the section *Data 'whitening'*.

An important feature of Eyenalysis is that it does not require an image to be segmented into RoIs. This is beneficial when such segmentation is difficult. But when RoIs are available, particularly semantically defined RoIs, this is a limitation. In such cases, a string edit algorithm is the method of choice. Among currently available string edit methods, ScanMatch (Cristino et al., 2010) is most sensitive, and should therefore be preferred over the classic Levenshtein distance (Levenshtein, 1966). Another useful feature of ScanMatch is that you can specify relationships between points in an image that violate geometric constraints (e.g., A→B > B →A), which is not possible in a geometric approach.

In summary, similarity measures are a powerful tool for eye movement research. We have proposed and validated a simple, yet sensitive algorithm for estimating the similarity between a pair of eye movement sequences.

∞

XII. CONCLUSION

My research project started with a simple question: Why do we perceive a stable world despite the continuously changing retinal image? Why do we not perceive a shift of the world with each eye movement? The research summarised in this thesis points towards a number of answers. These will be discussed in this final section, in which I will outline a general perspective on visual stability and discuss avenues for future research.

The subjective experience of visual stability does not require an explanation

Strikingly, one of the conclusions is that the very question that started the project is ill-posed. We do not consciously perceive the instability of the retinal image, because (trivially) we do not experience our retinal image. Conscious visual perception is an interpretation of visual input, and there is nothing contradictory about interpreting unstable input as reflecting a stable environment 'out there' (for similar ideas, see Deubel et al., 1998; MacKay, 1958, 1972; O'Regan & Noë, 2001). In other words, there is no need for an explicit mechanism that compensates for eye movements to explain why we consciously perceive the world as stable. When it comes to conscious visual perception, stability is an interpretation, or an assumption (see also *The assumption of stability* in Chapter ii).

It has taken me some time to acknowledge that what we were studying did not, in fact, have anything to do with conscious visual perception. But then—What have we been studying?

Visual stability is an issue related to attention and action

Based on the argument outlined above, some authors have dismissed visual stability as a non-issue altogether (e.g., O'Regan & Noë, 2001). I disagree. Even if we disregard the question of why we *perceive* a stable world, we are left with the question of why we *interact* with the world as though it is stable (Bays & Husain, 2007; see also Chapter v).

The present thesis is called *Visual Attention and Stability*. I believe that this title is fitting, because visual stability is essentially an issue related to attention and action. To illustrate this, let us consider what happens when you focus your attention on a coffee mug: You will perceive the mug more clearly in your mind's eye (Carrasco et al., 2000),

and you will prepare an action towards the coffee mug (presumably a reach-to-grasp and/ or an eye movement; Vainio, Ellis, & Tucker, 2007; see also Baldauf & Deubel, 2010; Deubel & Schneider, 1996).

From a neural perspective, when you shift your attention to the coffee mug there will be a localised change of activity throughout visual areas of the brain (for a review, see Colby & Goldberg, 1999). These brain areas are largely retinotopically organised, which means that neural activity reflects the location of a stimulus (the mug, in this case) on the retina, rather than the actual location of the object. Because of this retinotopic organisation, the neural activity that corresponds to the focus of attention will fall out of alignment with the world 'out there' when you make an eye movement (this is the classic 'problem' of visual stability, cf. Mathôt & Theeuwes, 2011a; Wurtz, 2008). The crucial question is, of course, whether such misalignment constitutes a problem. And as I have argued above, for many purposes it does not: The mug is right in front of your eyes, so after an eye movement you will receive new up-to-date retinal input. You don't need to update your internal representation of the mug to compensate for your eye movements, because you can rely on the physical mug as a gaze-invariant 'external representation' (O'Regan, 1992).

However, this line of reasoning does not apply to the focus of attention, which is an internal state that does not correspond to a physical property of the environment—There is no external representation to rely on. Yet the focus of attention is grounded in the visual system, and as such it is anchored to the direction of gaze. In other words, if there were no compensatory mechanism, the focus of attention would shift along with each eye movement, and we would be unable to maintain attention at the same location across eye movements. Crucially, because action preparation is a central aspect of visual attention (e.g., when you attend to a coffee mug, you automatically prepare a reaching movement), eye movements would disrupt action preparation.

Here then, we have a solid reason to posit an active mechanism that updates or 'remaps' information to compensate for eye movements. But this mechanism is not related to conscious visual perception. It serves only to update the focus of attention, and presumably other internal states like inhibition of return and working memory.

Remapping for visual stability is a peri-saccadic process

The temporal characteristics of remapping are known in considerable detail, in part because of the experiments described in this thesis (Chapters iii, iv, and v). Yet many questions remain.

In broad strokes, attentional effects that are commonly attributed to remapping are observed from around 150ms before to 150ms after a saccadic eye movement (Golomb et

al., 2008; Mathôt & Theeuwes, 2010a, 2010b; Rolfs et al., 2011). Immediately after a saccade, there are two distinct foci of attention. The first is a 'retinotopic trace' that corresponds to the retinal location of the focus of attention. Presumably, this retinotopic trace does not serve any purpose, but is an epiphenomenon resulting from the fact that attention is grounded in a retinotopically organised visual system. Consistent with this idea, the retinotopic trace dissipates quickly, within about 100 to 150ms. At the same time after a saccade, there is a separate focus of attention at the originally attended (spatiotopic) location. Although this spatiotopic focus of attention is already present immediately after a saccade (or at least as early as we can measure using psychophysical tools), there is a tendency for the spatiotopic focus of attention to grow stronger at longer intervals after a saccade. In other words, after a saccadic eye movement there is a brief, but measurable recovery period during which the (maladaptive) retinotopic focus of attention is suppressed and the (adaptive) spatiotopic focus of attention is consolidated (Golomb et al., 2008; Mathôt & Theeuwes, 2010b). We have interpreted this finding as reflecting an active remapping process that restores visual stability after a saccadic eye movement.

It is less clear what happens to the focus of attention just before a saccade. Whereas we showed a shift of attention in the direction of the impending saccade (Mathôt & Theeuwes, 2010a; see also W. Harrison et al., 2012), Rolfs and colleagues (2011) showed a shift in the opposite direction. Strikingly, these studies were based on the same neurophysiological assumptions, which illustrates that it is not straight-forward to derive predictions for behaviour from neurophysiological findings. In addition, based on a computational model as well as psychophysical data, Hamker and colleagues (2008; Zirnsak et al., 2011, 2010) argued that the focus of attention shifts in the direction of the saccade target, which essentially serves as an attractor (see Figure ii.4 for an illustration). Because the neural substrate of remapping is unknown, or at least debated (for different perspectives, see Cavanagh et al., 2010; Hamker et al., 2008; Mathôt & Theeuwes, 2011a; Morris et al., 2012; Wurtz, 2008), it is presently unclear whether these pre-saccadic effects are directly related to visual stability, and whether they should be viewed as functionally significant or as epiphenomena.

Finally, in experiments in which participants respond by making a saccadic eye movement, gaze dependence is much more pronounced for fast than for slow saccadic responses. Blohm and colleagues (2005) were the first to notice this, but to the best of my knowledge the experiments reported in Chapters iv and v are the only other studies that have explored this effect in detail (although see Overvliet et al., 2011 for similar results in the context of tactile perception). One interpretation is that fast saccadic eye movements

rely on a direct, subcortical pathway that bypasses the cortical brain areas involved in remapping for visual stability, but there is still much to be investigated.

Remapping for visual stability may be closely related to object-based attention

Remapping for visual stability is generally interpreted as a phenomenon that is specific to eye movements. But it may not be. Just as we are able to focus our attention on the same location across eye movements, we are able to focus our attention on the same spot of a moving object (Boi, Vergeer, Ogmen, & Herzog, 2011). In other words, the focus of attention matches the movement of an object, just like it compensates for the movement of our eyes. Clearly, these two situations are different in the sense that they are respectively related to actual and self-generated movement. But object-centred and world-centred (spatiotopic) attention might nevertheless rely on overlapping mechanisms. Indeed, recent preliminary studies from our lab suggest that this is the case (Boon, Theeuwes, & Belopolsky, in preparation; Theeuwes, Mathôt, & Grainger, in preparation).

An interesting avenue for future research will therefore be to investigate the link between object-based attention[17] and visual stability, two subjects which have traditionally been studied in isolation, but may not be so different.

Summary

We should not overstate the 'problem' of visual stability. There is no a-priori reason to believe, and no convincing data to suggest, that we consciously perceive a stable world because visual information is integrated from one fixation to the next. Instead, visual stability is an issue related to attention and action: Spatial constancy is preserved by an active remapping mechanism, but only insofar as required by our need to interact with the environment.

<p style="text-align:center">∞</p>

---

17 Object-based attention is generally studied by looking at the spreading of attention within the boundaries of an object (Egly, Driver, & Rafal, 1994; Roelfsema, Lamme, & Spekreijse, 1998; Theeuwes, Mathôt, & Kingstone, 2010). This is related, but not identical to the type of object-centred attention discussed here, in which the focus of attention is 'glued' to a moving object (Boi, Vergeer, Ogmen, & Herzog, 2011).

XIII. NEDERLANDSE SAMENVATTING

Visuele aandacht en stabiliteit

Mijn onderzoeksproject begon met een eenvoudige vraag: Hoe kan het dat we een stabiele wereld zien, ondanks het continu veranderende beeld op onze retina? Waarom zien we geen verschuiving van de wereld bij elke oogbeweging die we maken? In deze afsluitende sectie zal ik deze vraag beantwoorden op basis van het in mijn these besproken onderzoek. Daarnaast zal ik een aantal suggesties doen voor toekomstig onderzoek.

Het subjectieve gevoel van visuele stabiliteit behoeft geen verklaring

Eén van mijn belangrijkste conclusies is dat de bovengenoemde onderzoeksvraag feitelijk incorrect geformuleerd is. We zien de instabiliteit van het beeld op onze retina niet, simpelweg omdat we onze retina niet ervaren. Bewuste visuele perceptie is een *interpretatie* van het beeld dat op onze retina valt. En er is niets tegenstrijdigs aan het interpreteren van een instabiel retinaal beeld als veroorzaakt door een stabiele buitenwereld (voor vergelijkbare ideeën, zie Deubel et al., 1998; MackKay, 1958, 1972; O'Regan & Noë, 2001). Met andere woorden, er is geen expliciet, voor oogbewegingen corrigerend mechanisme nodig om te kunnen verklaren waarom we de wereld ervaren als stabiel. De verklaring is veel eenvoudiger: Als we het hebben over visuele perceptie dan is stabiliteit slechts een interpretatie, of een aanname.

Het heeft enige tijd geduurd voordat ik me besefte dat wat we aan het onderzoeken waren eigenlijk niets van doen had met bewuste visuele perceptie. Maar—Wat hebben we dan wel onderzocht?

Visuele stabiliteit heeft te maken met aandacht en actie

Op basis van de voorgaande redenering heeft een aantal wetenschappers visuele stabiliteit in zijn geheel afgedaan als een *non-issue* (b.v., O'Regan & Noë, 2001). Hier ben ik het echter niet mee eens. Zelfs als we ons niet afvragen hoe het kan dat we een stabiele wereld *zien*, blijft het de vraag waarom we *interacteren* met de wereld alsof deze stabiel is (Bays & Husain, 2007; zie ook Hoofdstuk v). Het antwoord ligt besloten in de titel van mijn these, *Visuele Aandacht en Stabiliteit*: Visuele stabiliteit is geen *non-issue* in situaties waarin aandacht en actie een rol spelen. Laten we ter illustratie eens kijken wat er gebeurt

wanneer je je aandacht richt op een koffiekopje. Het feit dat het kopje de focus van je aandacht is, zorgt ervoor dat je deze helderder in je geestesoog ziet (Carrasco et al., 2000), en dat je er automatisch een beweging naartoe voorbereidt (een grijpbeweging en/ of een oogbeweging; Vainio, Ellis, & Tucker, 2007; zie ook Baldauf & Deubel, 2010; Deubel & Schneider, 1996).

Op neuronaal niveau zorgt het richten van je aandacht op het kopje voor een locale verandering van activiteit in alle visuele hersengebieden (zie Colby & Goldberg, 1999 voor een overzichtsartikel). Deze hersengebieden zijn grotendeels retinotopisch georganiseerd. Dat wil zeggen dat neuronale activiteit de locatie van de stimulus (het kopje) op de retina weergeeft, en niet de daadwerkelijke locatie van het object in de buitenwereld. Als gevolg van deze retinotopische organisatie verschuift met elke oogbeweging de neuronale activiteit van de aandachtsfocus ten opzichte van de buitenwereld (het klassieke 'probleem' van visuele stabiliteit, cf. Mathôt & Theeuwes, 2011a; Wurtz, 2008). De cruciale vraag is uiteraard of deze verschuiving problematisch is. En, zoals ik hierboven heb beargumenteerd, is dit in veel gevallen niet zo: Het kopje blijft zichtbaar en zodoende ontvang je na elke oogbeweging nieuwe, actuele retinale input. De interne representatie van het kopje hoeft niet geactualiseerd te worden om te compenseren voor oogbewegingen, omdat het daadwerkelijke kopje gebruikt kan worden als een 'externe representatie' die niet gebonden is aan de positie van het oog (O'Regan, 1992).

Dezelfde redenering kan echter niet worden toegepast op de aandachtsfocus, omdat dit een interne toestand is en geen eigenschap van de omgeving—Er is geen externe representatie om op te vertrouwen. Desalniettemin is de aandachtsfocus geworteld in het visuele systeem en zodoende gebonden aan de positie van het oog. Met andere woorden, als er geen compensatiemechanisme zou bestaan dan zou de aandachtsfocus verschuiven met elke oogbeweging, en zouden we niet in staat zijn om onze aandacht op dezelfde locatie gevestigd te houden terwijl we oogbewegingen maken. Bovendien, gegeven de sterke relatie tussen aandacht en actiepreparatie (wanneer je je aandacht richt op een koffiekopje bereid je een beweging voor om deze op te pakken) zouden oogbewegingen een verstorend effect hebben op actiepreparatie.

En zo komen we tot een goede reden om aan te nemen dat er wel degelijk een actief compensatiemechanisme bestaat, niet ten behoeve van bewuste visuele perceptie, maar wel om interne toestanden zoals de aandachtsfocus (en waarschijnlijk ook *inhibition of return* en werkgeheugen) te updaten, of *remappen*.

Remapping voor visuele stabiliteit is een peri-saccadisch proces

De temporele eigenschappen van *remapping* (het mechanisme dat visuele stabiliteit ondersteunt) zijn, gedeeltelijk dankzij de in deze these beschreven experimenten (Hoofdstukken iii, iv en v), grotendeels bekend. Maar er zijn nog veel onbeantwoorde vragen.

Aan *remapping* toegeschreven aandachtseffecten doen zich met name voor van 150ms voor tot 150ms na een saccade van het oog (Golomb et al., 2008; Mathôt & Theeuwes, 2010a, 2010b; Rolfs et al., 2011). Direct na een saccade treden twee verschillende aandachtseffecten op. Ten eerste is de aandachtsfocus (of inhibitie) deels mee verschoven met het oog. Vermoedelijk heeft dit 'retinotopische spoor' geen functie, maar is het een epifenomeen dat voortkomt uit het feit dat aandacht is ingebed in een retinotopisch georganiseerd visueel systeem. Het retinotopische spoor verdwijnt dan ook zeer snel, binnen 100 tot 150ms. Tegelijkertijd ontstaat er na een saccade een tweede aandachtsfocus op de (spatiotopische) locatie waar de aandacht in eerste instantie op was gericht. Alhoewel deze spatiotopische aandachtsfocus zich al direct na de saccade voordoet (of in ieder geval zo snel als we kunnen meten met psychofysische technieken), wordt de focus sterker naarmate de tijd na de saccade verstrijkt. Met andere woorden, na een saccade van het oog is er een korte, maar meetbare, herstelperiode waarin de (maladaptieve) retinotopische aandachtsfocus wordt onderdrukt en de ('echte', adaptieve) spatiotopische aandachtsfocus wordt geconsolideerd. Wij hebben dit geïnterpreteerd als evidentie voor een actief *remappings*mechanisme dat visuele stabiliteit herstelt na een saccade.

Het is minder duidelijk wat er gebeurt met de aandachtsfocus net voor een saccade. Waar wij een verschuiving van aandacht in de richting van de aanstaande oogbeweging hebben aangetoond (Mathôt & Theeuwes, 2010a; zie ook W. Harrison et al., 2011), vonden Rolfs en collega's (2011) een verschuiving in de tegenovergestelde richting. Beide studies waren gebaseerd op dezelfde neurofysiologische aannamen, wat illustreert dat het niet eenvoudig is om gedragsvoorspellingen te doen aan de hand van neurofysiologische bevindingen. Verder hebben Hamker en collega's (2008; Zirnsak et al., 2011, 2010) op basis van zowel een computationeel model als psychofysische resultaten voorspeld dat de aandacht verschuift in de richting van het doel van de oogbeweging (zie Figuur ii.4 voor een illustratie). Omdat het neurale substraat van *remapping* onbekend of op zijn minst betwist is (voor diverse perspectieven zie Cavanagh et al., 2010; Hamker et al., 2008; Mathôt & Theeuwes, 2011a; Morris et al., 2012; Wurtz, 2008), blijft tot op heden onduidelijk of deze pre-saccadische effecten direct gerelateerd zijn aan visuele stabiliteit en of ze gezien moeten worden als functioneel of epifenomenaal.

Ten slotte, in experimenten waarbij proefpersonen reageren door het maken van een oogbeweging, zijn snelle reacties veel sterker afhankelijk van oogpositie dan langzame reacties. Blohm en collega's (2005) waren de eersten die dit opmerkten, maar voor zover ik weet zijn de in Hoofdstuk iv en v beschreven studies de enige andere die dit effect in detail hebben onderzocht (maar zie Overvliet et al., 2011 voor vergelijkbare resultaten voor tactiele perceptie). Eén mogelijke verklaring is dat snelle oogbewegingen hun grondslag hebben in een directe, sub-corticale route waarbij de bij *remapping* betrokken hersengebieden omzeild worden. Maar er valt nog veel te onderzoeken.

Remapping voor visuele stabiliteit is wellicht gerelateerd aan object-gebaseerde aandacht

Remapping voor visuele stabiliteit wordt vaak gezien als een fenomeen dat specifiek is voor oogbewegingen. Maar dit is wellicht niet het geval. We kunnen we onze aandacht op één positie binnen een bewegend object richten, net zoals we onze aandacht op dezelfde locatie gericht kunnen houden terwijl we oogbewegingen maken (Boi et al., 2011). Met andere woorden, de aandachtsfocus verschuift mee met een bewegend object, net zoals het compenseert voor een beweging van het oog. Natuurlijk zijn deze twee situaties verschillend, aangezien ze respectievelijk te maken hebben met 'echte' beweging en beweging die voortkomt uit onszelf. Desalniettemin zouden object- en wereld-georiënteerde (spatiotopische) aandacht op gedeeltelijk overlappende mechanismen kunnen berusten. Recente studies uit ons lab suggereren dat dit inderdaad het geval is (Boon, Theeuwes, & Belopolsky, in voorbereiding; Theeuwes, Mathôt, & Grainger, in voorbereiding).

Object-gebaseerde aandacht en visuele stabiliteit zijn grotendeels onafhankelijk van elkaar onderzocht.[18] De mogelijke relatie tussen beide fenomenen is zodoende een interessante richting voor toekomstig onderzoek.

Samenvatting

We moeten het 'probleem' van visuele stabiliteit niet overdrijven. Er is geen a-priori reden om te veronderstellen, en er zijn geen experimentele resultaten die suggereren, dat het feit dat we de wereld als stabiel ervaren voortkomt uit trans-saccadische integratie: Het beeld dat net voor een oogbeweging op je retina valt wordt nauwelijks geïntegreerd met

---

18 Object-gebaseerde aandacht wordt over het algemeen onderzocht door te kijken naar het verspreiden van aandacht binnen een object (Egly, Driver, & Rafal, 1994; Roelfsema, Lamme, & Spekreijse, 1998; Theeuwes, Mathôt, & Kingstone, 2010). Dit is gerelateerd aan, maar anders dan het type object-gebaseerde aandacht dat hier bedoeld wordt, waarbij de aandachtsfocus 'vast zit' aan een bewegend object (Bot et al., 2011).

het beeld dat na een oogbeweging op je retina valt. Visuele stabiliteit is een fenomeen dat te maken heeft met aandacht en actie: Visuele stabiliteit wordt ondersteund door een actief *remappings*mechanisme dat compenseert voor oogbewegingen, maar enkel in zoverre nodig is om effectief te interacteren met onze omgeving.

∞

XIV. GLOSSARY

Attention — There are many definitions of attention, but they generally resemble the one that I provide in the introduction: Attention is as a faculty of mind that allows us to focus on some aspect of our environment in order to see, hear, feel, or otherwise experience that particular aspect (more clearly).

Craniotopic (or: head-centred) — A frame of reference anchored to the position of the head; Relative to the head.

Eye-centred — See *retinotopic*.

Eye position signal — A pattern of neural activity that encodes the rotation of the eyes in the socket. An eye position signal has been recently identified in a number of dorsal visuomotor areas (Morris et al., 2012).

Gaze-centred — See *retinotopic*.

Head-centred — See *craniotopic*.

Non-representationalism — The conviction that visual experience does not arise from cognitive/ neural representations that are constructed based on sensory input, but is an immediate consequence of sensory input and active exploration. Noë and O'Regan describe (their variation of) this view as follows: "Instead of assuming that vision consists in the creation of an internal representation of the outside world whose activation somehow generates visual experience, we propose to treat vision as an exploratory activity" (O'Regan & Noë, 2001, p. 940).

Optokinetic reflex (OKR) — A type of reflexive eye movement that stabilises the retinal image based on visual input. The OKR 'glues' your eyes to the environment when you look out the window of a moving vehicle.

Persistent retinotopic bias — The phenomenon that our response is often biased by the retinal location of a stimulus, even when this does not match the actual location of a stimulus. An example of this, described in Chapter v, is the finding that we mislocalise briefly presented stimuli in the direction of a smooth pursuit eye movement.

Photoreceptors — Light sensitive cells in the retina.

Predictive remapping — The phenomenon, first demonstrated by Duhamel, Colby, and Goldberg (1992), that certain neurons appear to shift ('remap') their receptive field in the direction of an impending eye movement, just before the eyes actually set in motion. Predictive remapping is often linked to visual stability (but see Hamker et al., 2008; Morris et al., 2012; Zirnsak et al., 2010).

Remapping (or: spatial updating) — Without the prefix 'predictive', remapping is a contentious term. As used throughout most of this thesis, remapping refers to any mechanism that maintains visual stability by compensating for retinal displacements.

Representationalism — The conviction that visual experience arises from brain activity that somehow reflects the world 'out there'. Although rarely made explicit, representationalism is the dominant stream of thought within cognitive science (see e.g., Kanai & Tsuchiya, 2012).

Retina — The light sensitive layer containing photoreceptors in the back of the eye.

Retinal displacement — A shift of the retinal image relative to the world 'out there'. Saccades and smooth pursuit eye movements generally cause retinal displacements, whereas gaze stabilising reflexes prevent them.

Retinotopic (or: gaze-centred, eye-centred) — A frame of reference anchored to the retina; Relative to the eyes. Often contrasted with *spatiotopic*.

Retinotopic trace (or: transient retinotopic bias) — The phenomenon that immediately after an eye movement there is some transient retinotopic activity. This term was coined by Golomb and colleagues (2008) to describe their finding that just after an eye movement the locus of attention appears to have shifted along with the eye movement. See also Chapters iii and iv.

Saccade — A type of discrete voluntary eye movement that we use to scan our environment. Saccades shift our gaze from location to location in a shock-like fashion.

Smooth pursuit — A type of continuous voluntary eye movement that we use to track moving objects.

Spatial updating — See *remapping*.

Spatiotopic (or: world-centred) — A frame of reference anchored to the world 'out there'; Real position in space. Often contrasted with *retinotopic*. Because in most experiments

participants are not allowed to move their head, the term *spatiotopic* is often used loosely and cannot be distinguished from the weaker *craniotopic* frame of reference.

Stimulus onset asynchrony (SOA) — Technobabble for the interval between the presentation of two stimuli in an experiment.

Transient retinotopic bias — See *retinotopic trace*.

Vergence — A type of reflexive eye movement that centres both eyes on the same object, regardless of distance. For nearby objects this means that both eyes rotate inwards, for far away objects the eyes rotate (relatively) outwards.

Vestibulo-ocular reflex (VOR) — A type of reflexive eye movement that stabilises the retinal image through a counter-rotation of the eye when you make a head movement. The VOR allows you to shake your head while fixating.

Visual stability — The notion that visual perception does not appear to be affected by the instability that is characteristic of visual input at the level of the retina. In this thesis I distinguish between *visual stability for action* and *visual stability for perception* (cf. Bays & Husain, 2007).

Visual stability for action — The notion that we are able to execute visually guided movements, apparently unperturbed by the retinal displacements caused by eye movements. In line with the premotor theory of attention (Craighero & Rizzolatti, 2005), the ability maintain attention at a single location across eye movements is also a form of visual stability for action. See also Chapters v and vi.

Visual stability for perception — The notion that conscious visual experience is not affected by eye movements. See also Chapters v and vi.

Window of competition — The brief interval surrounding the presentation of a stimulus during which competitive interactions with other stimuli arise. See also Chapter viii.

Window of instability — The brief interval surrounding a saccadic eye movement during which visual stability breaks down and the retinotopic organisation of the visual system is most prominently observed using behavioural measures. See also Chapters iii and iv.

World-centred — See *spatiotopic*.

XV. BIBLIOGRAPHY

Abrams, R. A., & Dobkin, R. S. (1994). Inhibition of return: Effects of attentional cuing on eye movement latencies. *Journal of Experimental Psychology*, *20*(3), 467–477.

Abrams, R. A., & Pratt, J. (2000). Oculocentric coding of inhibited eye movements to recently attended locations. *Journal of Experimental Psychology: Human Perception and Performance*, *26*(2), 776–788.

Afraz, A., & Cavanagh, P. (2009). The gender-specific face aftereffect is based in retinotopic not spatiotopic across several natural image transformations. *Journal of Vision*, *9*(10), 1–17.

Awh, E., & Jonides, J. (1998). Spatial working memory and spatial selective attention. In R. Parasuraman (Ed.), *The Attentive Brain* (pp. 353–380). Cambridge, MA: MIT Press.

Awh, E., & Jonides, J. (2001). Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences*, *5*(3), 119–126.

Awh, E., Jonides, J., & Reuter-Lorenz, P. A. (1998). Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *24*(3), 780–790.

Azañón, E., Longo, M. R., Soto-Faraco, S., & Haggard, P. (2010). The posterior parietal cortex remaps touch into external Space. *Current Biology*, *20*, 1304–1309.

Azañón, E., & Soto-Faraco, S. (2008). Changing reference frames during the encoding of tactile events. *Current Biology*, *18*(14), 1044–1049. doi:10.1016/j.cub.2008.06.045

Bacon, J. H., Gordon, A., & Schulman, P. H. (1982). The effect of two types of induced-motion displays on perceived location of the induced target. *Attention, Perception, & Psychophysics*, *32*(4), 353–359.

Bahcall, D. O., & Kowler, E. (1999). Attentional interference at small spatial separations. *Vision Research*, *39*(1), 71–86.

Baldauf, D., & Deubel, H. (2010). Attentional landscapes in reaching and grasping. *Vision Research*, *50*(11), 999–1013.

Barash, S., Bracewell, R. M., Fogassi, L., Gnadt, J. W., & Andersen, R. A. (1991). Saccade-related activity in the lateral intraparietal area. II. Spatial properties. *Journal of Neurophysiology*, *66*(3), 1109–1124.

Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Perception & Psychophysics*, *28*(3), 241–248.

Bays, P. M., & Husain, M. (2007). Spatial remapping of the visual world across saccades. *Neuroreport*, *18*(12), 1207–1213. doi:10.1097/WNR.0b013e328244e6c3

Beck, D. M., & Kastner, S. (2005). Stimulus context modulates competition in human extrastriate cortex. *Nature Neuroscience*, *8*(8), 1110–1116.

Biber, U., & Ilg, U. J. (2011). Visual stability and the motion aftereffect: a psychophysical study revealing spatial updating. *PloS One*, *6*(1), e16265. doi:10.1371/journal.pone.0016265

Bisley, J. W., & Goldberg, M. E. (2010). Attention, Intention, and Priority in the Parietal Lobe. *Annual Review of Neuroscience*, *33*, 1–21. doi:10.1146/annurev-neuro-060909-152823

Blohm, G., Missal, M., & Lefevre, P. (2005). Processing of retinal and extraretinal signals for memory-guided saccades during smooth pursuit. *Journal of Neurophysiology*, *93*(3), 1510–1522.

Boi, M., Vergeer, M., Ogmen, H., & Herzog, M. H. (2011). Nonretinotopic Exogenous Attention. *Current Biology*. Retrieved from http://www.sciencedirect.com/science/article/pii/S0960982211009651

Boon, P., Theeuwes, J., & Belopolsky, A. V. (in preparation). *Updating Visual-spatial Memory During Object Movement*.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. doi:doi:10.1163/156856897X00357

Brand, A., & Bradley, M. T. (2011). Assessing the effects of technical variance on the statistical outcomes of web experiments measuring response times. *Social Science Computer Review*, *TODO*. doi:10.1177/0894439311415604

Brandt, S., & Stark, L. W. (1997). Spontaneous Eye Movements During Visual Imagery Reflect the Content of the Visual Scene. *Journal of Cognitive Neuroscience*, *9*(1), 27–38. doi:i: 10.1162/jocn.1997.9.1.27

Breitmeyer, B. G., Kropfl, W., & Julesz, B. (1982). The existence and role of retinotopic and spatiotopic forms of visual persistence. *Acta Psychologica*, *52*(3), 175–196.

Brenner, E., & Smeets, J. B. J. (1996). Size illusion influences how we lift but not how we grasp an object. *Experimental Brain Research*, *111*(3), 473–476. doi:10.1007/BF00228737

Brenner, E., & Smeets, J. B. J. (1997). Fast responses of the human hand to changes in target position. *Journal of Motor Behavior*, *29*(4), 297–310.

Brenner, E., Smeets, J. B. J., & van den Berg, A. V. (2001). Smooth eye movements and spatial localisation. *Vision Research*, *41*(17), 2253–2259. doi:10.1016/S0042-6989(01)00018-9

Bridgeman, B., Hendry, D., & Stark, L. (1975). Failure to detect displacement of the visual world during saccadic eye movements. *Vision Research*, *15*(6), 719–722.

Bridgeman, B., Kirch, M., & Sperling, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Attention, Perception, & Psychophysics*, *29*(4), 336–342. doi:10.3758/BF03207342

Bridgeman, B., & Mayer, M. (1983). Failure to integrate visual information from successive fixations. *Bulletin of the Psychonomic Society*, *21*(4), 285–286.

Brockmole, J. R., & Henderson, J. M. (2005a). Prioritization of new objects in real-world scenes: evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(5), 857–868.

Brockmole, J. R., & Henderson, J. M. (2005b). Object appearance, disappearance, and attention prioritization in real-world scenes. *Psychonomic Bulletin & Review*, *12*(6), 1061–1067.

Burr, D. C., & Ross, J. (1982). Contrast sensitivity at high velocities. *Vision research*, *22*(4), 479–484.

Caldara, R., & Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods*, *43*(3), 864–878. doi:10.3758/s13428-011-0092-x

Cameron, B. D., Enns, J. T., Franks, I. M., & Chua, R. (2009). The hand's automatic pilot can update visual information while the eye is in motion. *Experimental Brain Research*, *195*(3), 445–454.

Campbell, F. W., & Wurtz, R. H. (1978). Saccadic omission: why we do not see a grey-out during a saccadic eye movement. *Vision Research*, *18*(10), 1297–1303.

Caputo, G., & Guerra, S. (1998). Attentional selection by distractor suppression. *Vision Research*, *38*(5), 669–689.

Carrasco, M. (2011). Visual attention. *Vision Research*, *51*(13), 1484–1525. doi:10.1016/j.visres.2011.04.012

Carrasco, M., Penpeci-Talgar, C., & Eckstein, M. (2000). Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision Research*, *40*, 1203–1215. doi:10.1016/S0042-6989(00)00024-9

Castet, E. (2010). Perception of intra-saccadic motion. In U. J. Ilg & G. S. Masson (Eds.), *Dynamics of Visual Motion Processing* (pp. 213–238). Springer Science.

Castet, E., & Masson, G. S. (2000). Motion perception during saccadic eye movements. *Nature Neuroscience*, *3*(2), 177–183.

Cavanagh, P., Hunt, A. R., Afraz, A., & Rolfs, M. (2010). Visual stability based on remapping of attention pointers. *Trends in Cognitive Sciences*, *14*(4), 147–153.

Cave, K. R., & Zimmerman, J. M. (1997). Flexibility in spatial attention before and after practice. *Psychological Science*, *8*(5), 399–403.

Cha, O., & Chong, S. C. (2010). Background is remapped across saccades. *Journal of Vision*, *10*(7), 516–516.

doi:10.1167/10.7.516

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940.

Chua, F. K. (2009). A new object captures attention—but only when you know it's new. *Attention, Perception, & Psychophysics*, *71*(4), 699–711. doi:10.3758/APP.71.4.699

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: An interactive graphic system for designing and controlling experiments in the psychology laboratory using Macintosh computers. *Behavior Research Methods: Instruments & Computers*, *25*(2), 257–271.

Colby, C. L. (1996). A neurophysiological distinction between attention and intention. In T. Inui, D. Gopher, A. L. McClelland, & A. Koriat (Eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 157–181). Cambridge, MA: MIT Press.

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1996). Visual, presaccadic, and cognitive activation of single neurons in monkey lateral intraparietal area. *Journal of Neurophysiology*, *76*(5), 2841–2852.

Colby, C. L., & Goldberg, M. E. (1999). Space and attention in parietal cortex. *Annual Review of Neuroscience*, *22*, 319–349.

Connor, C. E., Gallant, J. L., Preddie, D. C., & Van Essen, D. C. (1996). Responses in area V4 depend on the spatial relationship between stimulus and attention. *Journal of Neurophysiology*, *75*(3), 1306–1308.

Connor, C. E., Preddie, D. C., Gallant, J. L., & Van Essen, D. C. (1997). Spatial attention effects in macaque area V4. *Journal of Neuroscience*, *17*(9), 3201–3214.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology*, *1*(1), 4–45.

Craighero, L., & Rizzolatti, G. (2005). The premotor theory of attention. In L. Itti, G. Rees, & J. K. Tsotsos (Eds.), *Neurobiology of Attention* (pp. 181–186). Amsterdam, Netherlands: Elsevier.

Crespi, S., Biagi, L., d' Avossa, G., Burr, D. C., Tosetti, M., & Morrone, M. C. (2011). Spatiotopic coding of BOLD signal in human visual cortex depends on spatial attention. *PLoS ONE*, *6*(7), e21661. doi:10.1371/journal.pone.0021661

Cristino, F., Mathôt, S., Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692–700. doi:10.3758/BRM.42.3.692

Currie, C. B., McConkie, G. W., Carlson-Radvansky, L. A., & Irwin, D. E. (2000). The role of the saccade target object in the perception of a visually stable world. *Perception & Psychophysics*, *62*(4), 673–683.

Damian, M. F. (2010). Does variability in human performance outweigh imprecision in response devices such as computer keyboards? *Behavior Research Methods*, *42*, 205–211. doi:10.3758/BRM.42.1.205

Davoli, C. C., Suszko, J. W., & Abrams, R. A. (2007). New objects can capture attention without a unique luminance transient. *Psychonomic Bulletin & Review*, *14*(2), 338–343. doi:10.3758/BF03194074

d' Avossa, G., Tosetti, M., Crespi, S., Biagi, L., Burr, D. C., & Morrone, M. C. (2007). Spatiotopic selectivity of BOLD responses to visual motion in human area MT. *Nature Neuroscience*, *10*(2), 249–255.

de Hoon, M., Imoto, S., & Miyano, S. (2010). *The C Clustering library*. Retrieved from http://bonsai.hgc.jp/~mdehoon/software/cluster/software.htm

Demeyer, M., De Graef, P., Wagemans, J., & Verfaillie, K. (2009). Transsaccadic identification of highly similar artificial shapes. *Journal of Vision*, *9*(4), 1–14. doi:10.1167/9.4.28

Demeyer, M., De Graef, P., Wagemans, J., & Verfaillie, K. (2010). Parametric integration of visual form across saccades. *Vision Research*, *50*(13), 1225–1234. doi:10.1016/j.visres.2010.04.008

Dempere-Marco, L., Hu, X. P., Ellis, S. M., Hansell, D. M., & Yang, G. Z. (2006). Analysis of visual search patterns with EMD metric in normalized anatomical space. *IEEE Transactions on Medical Imaging*, *25*(8), 1011–1021.

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical*

Transactions of the Royal Society B: Biological Sciences, *353*(1373), 1245–1255.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Desimone, R., & Schein, S. J. (1987). Visual properties of neurons in area V4 of the macaque: sensitivity to stimulus form. *Journal of Neurophysiology*, *57*(3), 835–868.

Deubel, H. (2004). Localization of targets across saccades: Role of landmark objects. *Visual Cognition*, *11*(2), 173–202.

Deubel, H., Bridgeman, B., & Schneider, W. X. (1998). Immediate post-saccadic information mediates space constancy. *Vision Research*, *38*(20), 3147–3159.

Deubel, H., Irwin, D. E., & Schneider, W. X. (1999). The subjective direction of gaze shifts long before the saccade. In W. Becker, H. Deubel, & T. Mergner (Eds.), *Current Oculomotor Research: Physiological and Psychological Aspects* (pp. 65–70). New York, NY: Kluwer Academic.

Deubel, H., Koch, C., & Bridgeman, B. (2010). Landmarks facilitate visual space constancy across saccades and during fixation. *Vision Research*, *50*(2), 249–259.

Deubel, H., & Schneider, W. X. (1994). Perceptual stability and postsaccadic visual information: Can man bridge a gap? *Behavioral and Brain Sciences*, *17*(2), 259–259.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

Deubel, H., Schneider, W. X., & Bridgeman, B. (1996). Postsaccadic target blanking prevents saccadic suppression of image displacement. *Vision Research*, *36*(7), 985–996.

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science*, *6*(3), 274–290. doi:10.1177/1745691611406920

Di Lollo, V. (1980). Temporal integration in visual memory. *Journal of Experimental Psychology: General*, *109*(1), 75–97.

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431.

Donk, M., & van Zoest, W. (2008). Effects of salience are short-lived. *Psychological Science*, *19*(7), 733–739.

Downing, C. J. (1988). Expectancy and visual-spatial attention: Effects on perceptual quality. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(2), 188–202.

Downing, C. J., Pinker, S., Posner, M. I., & Marin, O. S. M. (1985). The spatial structure of visual attention. *Attention and Performance XI* (pp. 171–187). Hillsdale, NJ: Erlbaum.

Duchowski, A. T., Driver, J., Jolaoso, S., Tan, W., Ramey, B. N., & Robbins, A. (2010). Scanpath comparison revisited. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 219–226).

Duhamel, J. R., Bremmer, F., BenHamed, S., & Graf, W. (1997). Spatial invariance of visual receptive fields in parietal cortex neurons. *Nature*, *389*, 845–848.

Duhamel, J. R., Colby, C. L., & Goldberg, M. E. (1992). The updating of the representation of visual space in parietal cortex by intended eye movements. *Science*, *255*(5040), 90–92.

Duncker, K. (1929). Über induzierte bewegung. *Psychologischen Forschung*, *12*(1), 180–259.

Dyde, R., & Milner, A. (2002). Two illusions of perceived orientation: one fools all of the people some of the time; the other fools all of the people all of the time. *Experimental Brain Research*, *144*(4), 518–527. doi:10.1007/s00221-002-1065-1

Easter, S. S. (1972). Pursuit eye movements in goldfish (Carassius auratus). *Vision research*, *12*(4), 673–688.

Eaton, J. W. (2002). *GNU Octave Manual*. Network Theory Limited.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General*, *123*(2), 161–177.

Ezzati, A., Golzar, A., & Afraz, A. S. R. (2008). Topography of the motion aftereffect with and without eye. *Journal of Vision*, *8*(14), 1–16.

Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (2000). Visuomotor neurons: ambiguity of the discharge or

"motor" perception? *International Journal of Psychophysiology*, *35*, 165–177.

Ferman, L., Collewijn, H., Jansen, T. C., & Van den Berg, A. V. (1987). Human gaze stability in the horizontal, vertical and torsional direction during voluntary head movements, evaluated with a three-dimensional scleral induction coil technique. *Vision Research, 27*(5), 811–828.

Filehne, W. (1922). Über das optische Wahrnehmen von Bewegungen. *Zeitschrift für Sinnephysiologie*, *53*, 134–145.

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

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods: Instruments, & Computers*, *35*(1), 116–124.

Foulsham, T., & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, *8*(2), 1–17.

Franconeri, S. L., Hollingworth, A., & Simons, D. J. (2005). Do new objects capture attention? *Psychological Science*, *16*(4), 275–281.

Franconeri, S. L., & Simons, D. J. (2003a). Moving and looming stimuli capture attention. *Perception & Psychophysics*, *65*(7), 999–1010. doi:10.3758/BF03194829

Franconeri, S. L., & Simons, D. J. (2003b). Moving and looming stimuli capture attention. *Attention, Perception, & Psychophysics*, *65*(7), 999–1010.

Galletti, C., Battaglini, P. P., & Fattori, P. (1993). Parietal neurons encoding spatial locations in craniotopic coordinates. *Experimental Brain Research*, *96*(2), 221–229.

Galletti, C., Battaglini, P. P., & Fattori, P. (1995). Eye position influence on the parieto-occipital area PO (V6) of the macaque monkey. *European Journal of Neuroscience*, *7*(12), 2486–2501.

Gardner, J. L., Merriam, E. P., Movshon, J. A., & Heeger, D. J. (2008). Maps of visual space in human occipital cortex are retinotopic, not spatiotopic. *Journal of Neuroscience*, *28*(15), 3988–3999. doi:10.1523/JNEUROSCI.5476-07.2008

Gattass, R., Sousa, A. P., & Gross, C. G. (1988). Visuotopic organization and extent of V3 and V4 of the macaque. *Journal of Neuroscience*, *8*(6), 1831–1845.

Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: A cross-platform experiment-programming library. *Behavior Research Methods*, *39*(4), 950–958.

Gellman, R. S., & Fletcher, W. A. (1992). Eye position signals in human saccadic processing. *Experimental Brain Research*, *89*(2), 425–434. doi:10.1007/BF00228258

Gerrits, H. J. M., De Haan, B., & Vendrik, A. J. H. (1966). Experiments with retinal stabilized images. Relations between the observations and neural data. *Vision Research*, *6*, 427–440. doi:10.1016/0042-6989(66)90051-4

Gibboni, R. R., Zimmerman, P. E., & Gothard, K. M. (2009). Individual differences in scanpaths correspond with serotonin transporter genotype and behavioral phenotype in Rhesus monkeys (Macaca mulatta). *Frontiers in Behavioral Neuroscience*, *3*, e50. doi:10.3389/neuro.08.050.2009

Gibson, J. J., & Radner, M. (1937). Adaptation, after-effect and contrast in the perception of tilted lines. I. Quantitative studies. *Journal of Experimental Psychology*, *20*(5), 453–467.

Godijn, R., & Pratt, J. (2002). Endogenous saccades are preceded by shifts of visual attention: Evidence from cross-saccadic priming effects. *Acta Psychologica, 110*(1), 83–102.

Goldberg, M. E., & Bruce, C. J. (1990). Primate frontal eye fields. III. Maintenance of a spatially accurate saccade signal. *Journal of Neurophysiology*, *64*(2), 489–508.

Golomb, J. D., Chun, M. M., & Mazer, J. A. (2008). The native coordinate system of spatial attention is retinotopic. *Journal of Neuroscience*, *28*(42), 10654–10662.

Golomb, J. D., & Kanwisher, N. (in press). Higher level visual cortex represents retinotopic, not spatiotopic, object location. *Cerebral Cortex*. doi:10.1093/cercor/bhr357

Golomb, J. D., & Kanwisher, N. (2012). Retinotopic memory is more precise than spatiotopic memory. *Proceedings of*

the *National Academy of Sciences*, *109*(5), 1796–1801.

Golomb, J. D., Marino, A. C., Chun, M. M., & Mazer, J. A. (2011). Attention doesn't slide: spatiotopic updating after eye movements instantiates a new, discrete attentional locus. *Attention, Perception, & Psychophysics*, *73*, 1–8.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20.

Gottlieb, J. P., Kusunoki, M., & Goldberg, M. E. (1998). The representationof visual salience in monkey parietal cortex. *Nature*, *391*, 481–484.

Green, M. (1991). Visual search, visual streams, and visual architectures. *Attention, Perception, & Psychophysics*, *50*(4), 388–403.

Green, M. (1992). Visual search: Detection, identification, and localization. *Perception*, *21*, 765–765.

Grimes, J. (1996). On the failure to detect changes in scenes across saccades. *Perception*, *5*, 89–110.

Grindinger, T., Murali, V., Tetreault, S., Duchowski, A., Birchfield, S., & Orero, P. (2011). Algorithm for discriminating aggregate gaze points: comparison with salient regions-of-interest. *Computer Vision–ACCV 2010 Workshops* (pp. 390–399).

Grüsser, O. J. (1986). Interaction of efferent and afferent signals in visual perception A history of ideas and experimental paradigms. *Acta Psychologica*, *63*(1), 3–21.

Gysen, V., De Graef, P., & Verfaillie, K. (2002). Detection of intrasaccadic displacements and depth rotations of moving objects. *Vision Research*, *42*(3), 379–391.

Hacisalihzade, S. S., Stark, L. W., & Allen, J. S. (1992). Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Transactions on systems, man, and cybernetics*, *22*(3), 474–481.

Hamker, F. H., Zirnsak, M., Calow, D., & Lappe, M. (2008). The peri-saccadic perception of objects and space. *PLoS Computational Biology*, *4*(2), e31.

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. doi:10.1038/nature07832

Harrison, W., Remington, R., & Mattingley, J. (2012). Pre-saccadic shifts of visual attention. *PLoS ONE*, *7*(9), e45670. doi:10.1167/11.11.242

Hartmann, T. S., Bremmer, F., Albright, T. D., & Krekelberg, B. (2011). Receptive field positions in area MT during slow eye movements. *The Journal of Neuroscience*, *31*(29), 10437 –10444. doi:10.1523/JNEUROSCI.5590-10.2011

Hazelhoff, F. F., & Wiersma, H. (1924). Die Wahrnehmungszeit. *Zeitschrift für Psychologie*, *96*, 171–188.

Henderson, J. M., Brockmole, J. R., Castelhano, M. S., & Mack, M. (2007). Visual saliency does not account for eye movements during visual search in real-world scenes. In M. Fischer, W. Murray, R. Hill, & R. P. G. van Gompel (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 537–562). Amsterdam, Netherlands: Elsevier.

Henderson, J. M., Pollatsek, A., & Rayner, K. (1987). Effects of foveal priming and extrafoveal preview on object identification. *Journal of Experimental Psychology: Human Perception and Performance*, *13*(3), 449–463.

Henderson, J. M., & Siefert, A. B. C. (2001). Types and tokens in transsaccadic object identification: Effects of spatial position and left-right orientation. *Psychonomic Bulletin & Review*, *8*(4), 753–760.

Heywood, S., & Churcher, J. (1981). Direction-specific and position-specific effects upon detection of displacements during saccadic eye movements. *Vision Research*, *21*(2), 255–261.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135–145.

Hilchey, M. D., Klein, R. M., Satel, J., & Wang, Z. (in press). Oculomotor inhibition of return: How soon is it "recoded" into spatiotopic coordinates? *Attention, Perception, & Psychophysics*.

Hoffman, J. E., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics*, *57*(6), 787–795.

Hollingworth, A., Richard, A. M., & Luck, S. J. (2008). Understanding the function of visual short-term memory: Transsaccadic memory, object correspondence, and gaze correction. *Journal of Experimental Psychology: General*, *137*(1), 163–181.

Hollingworth, A., Simons, D. J., & Franconeri, S. L. (2010). New objects do not capture attention without a sensory transient. *Attention, Perception & Psychophysics*, *72*(5), 1298–1310. doi:10.3758/APP.72.5.1298

Holt, E. B. (1903). Eye-movement and central anaesthesia. *Psychological Monographs*, *4*(1), 1–45.

Hopf, J. M., Boehler, C. N., Luck, S. J., Tsotsos, J. K., Heinze, H. J., & Schoenfeld, M. A. (2006). Direct neurophysiological evidence for spatial suppression surrounding the focus of attention in vision. *Proceedings of the National Academy of Sciences*, *103*(4), 1053–1058.

Hughes, H. C., & Zimba, L. D. (1985). Spatial maps of directed visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(4), 409–430.

Hughes, H. C., & Zimba, L. D. (1987). Natural boundaries for the spatial spread of directed visual attention. *Neuropsychologia*, *25*(1A), 5–18.

Hunt, A. R., & Cavanagh, P. (2009). Looking ahead: The perceived direction of gaze shifts before the eyes move. *Journal of Vision*, *9*(9), 1–7.

Hunt, A. R., & Cavanagh, P. (2011). Remapped visual masking. *Journal of Vision*, *11*(1), e13. doi:10.1167/11.1.13

Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognitive Psychology*, *23*(3), 420–456.

Irwin, D. E. (1992a). Perceiving an integrated visual world. In D. E. Meyer (Ed.), *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience* (pp. 121–142). Cambridge, MA: MIT Press.

Irwin, D. E. (1992b). Memory for position and identity across eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 307–317.

Irwin, D. E. (1996). Integrating information across saccadic eye movements. *Current Directions in Psychological Science*, *5*(3), 94–100.

Irwin, D. E., & Andrews, R. (1996). Integration and accumulation of information across saccadic eye movements. In T. Inui, D. Gopher, A. L. McClelland, & A. Koriat (Eds.), *Attention and Performance XVI: Information Integration in Perception and Communication* (pp. 125–155). Cambridge, MA: MIT Press.

Irwin, D. E., & Gordon, R. D. (1998). Eye movements, attention and trans-saccadic memory. *Visual Cognition*, *5*, 127–155.

Irwin, D. E., Yantis, S., & Jonides, J. (1983). Evidence against visual integration across saccadic eye movements. *Perception & Psychophysics*, *34*(1), 49–57.

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

Itti, L., & Koch, C. (2001). Computational modelling of visual attention. *Nature Reviews Neuroscience*, *2*(3), 194–204.

Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.

James, W. (1890). *The Principles of Psychology* (Vol. 1). Boston, MA: Harvard University Press.

Jarodzka, H., Holmqvist, K., & Nyström, M. (2010). A vector-based, multidimensional scanpath similarity measure. *Proceedings of the 2010 Symposium on Eye Tracking* (pp. 211–218). ACM Press. doi:10.1145/1743666.1743718

Jeffreys, H. (1961). *Theory of Probability*. Oxford, UK: Oxford University Press.

Jones, E., Oliphant, T., & Peterson, P. (2001). *Scipy: Open source scientific tools for Python*. Retrieved from http://www.scipy.org/

Jonides, J., Irwin, D. E., & Yantis, S. (1982). Integrating visual information from successive fixations. *Science*, *215*(4529), 192–194.

Jonides, J., & Yantis, S. (1988). Uniqueness of abrupt visual in capturing attention. *Perception & Psychophysics*, *43*(4),

346–354.

Kahneman, D., Treisman, A., & Gibbs, B. J. (1992). The reviewing of object files: Object-specific integration of information. *Cognitive psychology*, *24*(2), 175–219.

Kaminiarz, A., Krekelberg, B., & Bremmer, F. (2007). Localization of visual targets during optokinetic eye movements. *Vision Research*, *47*(6), 869–878. doi:10.1016/j.visres.2006.10.015

Kanai, R., & Tsuchiya, N. (2012). Qualia. *Current Biology*, *22*(10), 392–396. doi:10.1016/j.cub.2012.03.033

Kastner, S., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1998). Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science*, *282*(5386), 108–111.

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, *86*(3), 1398–1411.

Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, *69*(10), 1340–1349.

Keysers, C., & Perrett, D. I. (2002). Visual masking and RSVP reveal neural competition. *Trends in Cognitive Sciences*, *6*(3), 120–125.

Khayat, P. S., Spekreijse, H., & Roelfsema, P. R. (2004). Correlates of transsaccadic integration in the primary visual cortex of the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(34), 12712–12717.

Klein, R. M. (1988). Inhibitory tagging system facilitates visual search. *Nature*, *334*, 430–431.

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147.

Klein, R. M., & MacInnes, W. J. (1999). Inhibition of return Is a foraging facilitator in visual search. *Psychological Science*, *10*(4), 346–352.

Knapen, T., Rolfs, M., & Cavanagh, P. (2009). The reference frame of the motion aftereffect is retinotopic. *Journal of Vision*, *9*(5), 1–7. doi:10.1167/9.5.16

Knapen, T., Rolfs, M., Wexler, M., & Cavanagh, P. (2010). The reference frame of the tilt aftereffect. *Journal of Vision*, *10*(1), 1–13.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, *4*(4), 219–27.

Kowler, E. (2011). Eye movements: The past 25 years. *Vision Research*, *51*(13), 1457–1483. doi:16/j.visres.2010.12.014

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research*, *35*(13), 1897–1916.

Krauzlis, R. J., & Stone, L. S. (1999). Tracking with the mind's eye. *Trends in neurosciences*, *22*(12), 544–550.

Kuang, X., Poletti, M., Victor, J. D., & Rucci, M. (2012). Temporal encoding of spatial information during active visual fixation. *Current Biology*, *22*(6), 510–514. doi:10.1016/j.cub.2012.01.050

Kusunoki, M., & Goldberg, M. E. (2003). The time course of perisaccadic receptive field shifts in the lateral intraparietal area of the monkey. *Journal of Neurophysiology*, *89*(3), 1519–1527.

Land, M. F. (1999). Motion and vision: why animals move their eyes. *Journal of Comparative Physiology A: Neuroethology, Sensory, Neural, and Behavioral Physiology*, *185*(4), 341–352.

Land, M. F., & Nilsson, D. E. (2002). *Animal eyes*. Oxford, UK: Oxford University Press.

Levenshtein, V. I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, *10*(8), 707–710.

Logothetis, N. K., & Schall, J. D. (1989). Neuronal correlates of subjective visual perception. *Science*, *245*(4919), 761–763.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of Neurophysiology*, *77*(1), 24–42.

Luck, S. J., Girelli, M., McDermott, M. T., & Ford, M. A. (1997). Bridging the gap between monkey neurophysiology and human perception: An ambiguity resolution theory of visual selective attention. *Cognitive Psychology*, *33*(1), 64–87.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281.

Ludwig, C. J. H., Farrell, S., Ellis, L. A., & Gilchrist, I. D. (2009). The mechanism underlying inhibition of saccadic return. *Cognitive Psychology*, *59*(2), 180–202. doi:10.1016/j.cogpsych.2009.04.002

Mack, A. (2003). Inattentional blindness: looking without seeing. *Current Directions in Psychological Science*, *12*(5), 180–184.

Mack, A., & Herman, E. (1973). Position constancy during pursuit eye movement: An investigation of the filehne illusion. *Quarterly Journal of Experimental Psychology*, *25*(1), 71–84. doi:10.1080/14640747308400324

Mack, A., & Rock, I. (1998). *Inattentional blindness: Perception without attention*. Cambridge, MA: MIT Press.

MacKay, D. M. (1958). Perceptual Stability of a Stroboscopically Lit Visual Field containing Self-Luminous Objects. *Nature*, *181*(4607), 507–508. doi:10.1038/181507a0

MacKay, D. M. (1972). Visual stability. *Investigative Ophthalmology*, *11*, 518–524.

Mangun, G. R., & Hillyard, S. A. (1988). Spatial gradients of visual attention: Behavioral and electrophysiological evidence. *Electroencephalography and Clinical Neurophysiology*, *70*(5), 417–428.

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

Mannan, S. K., Ruddock, K. H., & Wooding, D. S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, *26*, 1059–1072.

Martinez-Conde, S., & Macknik, S. L. (2008). Fixational eye movements across vertebrates: Comparative dynamics, physiology, and perception. *Journal of Vision*, *8*(14), e28. doi:10.1167/8.14.28

Mateeff, S., Mitrani, L., & Stojanova, J. (1982). Visual localization during eye tracking on steady background and during steady fixation on moving background. *Biological Cybernetics*, *42*(3), 215–219. doi:10.1007/BF00340078

Mateeff, S., Yakimoff, N., & Dimitrov, G. (1981). Localization of brief visual stimuli during pursuit eye movements. *Acta Psychologica*, *48*(1–3), 133–140. doi:10.1016/0001-6918(81)90055-X

Mathôt, S., Cristino, F., Gilchrist, I. D., & Theeuwes, J. (2012). A simple way to estimate similarity between pairs of eye movement sequences. *Journal of Eye Movement Research*, *5*(1), 1–15.

Mathôt, S., Hickey, C., & Theeuwes, J. (2010). From reorienting of attention to biased competition: Evidence from hemifield effects. *Attention, Perception & Psychophysics*, *72*(3), 651–657.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314–324.

Mathôt, S., & Theeuwes, J. (2010a). Evidence for the predictive remapping of visual attention. *Experimental Brain Research*, *200*(1), 117–122. doi:10.1007/s00221-009-2055-3

Mathôt, S., & Theeuwes, J. (2010b). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychological Science*, *21*(12), 1793–1798.

Mathôt, S., & Theeuwes, J. (2011a). Visual attention and stability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*, 516–527.

Mathôt, S., & Theeuwes, J. (2011b). Mantra: An open method for object and movement tracking. *Behavior Research Methods*, *43*(4), 1182–1193.

Mathôt, S., & Theeuwes, J. (2012). It's all about the transient: Intra-saccadic abrupt onsets do not capture attention. *Journal of Eye Movement Research*, *5*(2), 1–12.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin*, *81*(12), 899–917.

Matsumiya, K., & Uchikawa, K. (2000). Distortion of visual space during pursuit eye movements. *Optical Review*,

7(3), 241–248. doi:10.1007/s10043-000-0241-4

Maylor, E. A., & Hockey, R. (1985). Inhibitory component of externally controlled covert orienting in visual space. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(6), 777–787.

McConkie, G. W., & Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(3), 563–581.

McConkie, G. W., & Rayner, K. (1976). Identifying the span of the effective stimulus in reading: Literature review and theories of reading. In H. Singer & R. B. Ruddell (Eds.), *Theoretical Models and Processes of Reading* (pp. 137–162). Newark, DE: International Reading Association.

McConkie, G. W., & Zola, D. (1979). Is visual information integrated across successive fixations in reading? *Perception & Psychophysics*, *25*(3), 221–224.

McKenzie, A., & Lisberger, S. G. (1986). Properties of signals that determine the amplitude and direction of saccadic eye movements in monkeys. *Journal of Neurophysiology*, *56*(1), 196–207.

McKyton, A., Pertzov, Y., & Zohary, E. (2009). Pattern matching is assessed in retinotopic coordinates. *Journal of Vision*, *9*(13), e19. doi:10.1167/9.13.19

Medendorp, W. P., Goltz, H. C., Vilis, T., & Crawford, J. D. (2003). Gaze-centered updating of visual space in human parietal cortex. *Journal of Neuroscience*, *23*(15), 6209–6214.

Melcher, D. (2005). Spatiotopic transfer of visual-form adaptation across saccadic eye movements. *Current Biology*, *15*(19), 1745–1748.

Melcher, D. (2007). Predictive remapping of visual features precedes saccadic eye movements. *Nature Neuroscience*, *10*(7), 903–907.

Melcher, D. (2008a). Dynamic, object-based remapping of visual features in trans-saccadic perception. *Journal of Vision*, *8*(14), 1–17. doi:10.1167/8.14.2

Melcher, D. (2008b). Selective attention and the active remapping of object features in trans-saccadic perception. *Vision Research*, *49*(10), 1249–1255.

Melcher, D., & Fracasso, A. (2012). Remapping of the line motion illusion across eye movements. *Experimental Brain Research*, *218*(4), 503–514. doi:10.1007/s00221-012-3043-6

Melcher, D., & Morrone, M. C. (2003). Spatiotopic temporal integration of visual motion across saccadic eye movements. *Nature Neuroscience*, *6*(8), 877–881.

Merriam, E. P., & Colby, C. L. (2005). Active vision in parietal and extrastriate cortex. *The Neuroscientist*, *11*(5), 484–493.

Merriam, E. P., Genovese, C. R., & Colby, C. L. (2003). Spatial updating in human parietal cortex. *Neuron*, *39*(2), 361–373.

Merriam, E. P., Genovese, C. R., & Colby, C. L. (2007). Remapping in human visual cortex. *Journal of Neurophysiology*, *97*(2), 1738–1755.

Milner, A. D. (1998). Streams and consciousness: visual awareness and the brain. *Trends in Cognitive Sciences*, *2*(1), 25–30.

Mishkin, M., Ungerleider, L. G., & Macko, K. A. (1983). Object vision and spatial vision: Two cortical pathways. *Trends in Neurosciences*, *6*(10), 414–417.

Mitrani, L., & Dimitrov, G. (1982). Retinal location and visual localization during pursuit eye movement. *Vision Research*, *22*(8), 1047–1051. doi:10.1016/0042-6989(82)90041-4

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science*, *229*(4715), 782–784.

Morris, A. P., Chambers, C. D., & Mattingley, J. B. (2007). Parietal stimulation destabilizes spatial updating across saccadic eye movements. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(21), 9069–9074.

Morris, A. P., Kubischik, M., Hoffmann, K.-P., Krekelberg, B., & Bremmer, F. (2012). Dynamics of eye-position

signals in the dorsal visual system. *Current Biology*, *22*(3), 173–179. doi:10.1016/j.cub.2011.12.032

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of Neurophysiology*, *70*(3), 909–919.

Mounts, J. R. W. (2000a). Attentional capture by abrupt onsets and feature singletons produces inhibitory surrounds. *Perception & Psychophysics*, *62*(7), 1485–1493.

Mounts, J. R. W. (2000b). Evidence for suppressive mechanisms in attentional selection: Feature singletons produce inhibitory surrounds. *Perception & Psychophysics*, *62*(5), 969–983.

Mounts, J. R. W., & Gavett, B. E. (2004). The role of salience in localized attentional interference. *Vision Research*, *44*(13), 1575–1588.

Mueller, S. T. (2010). *The PEBL Manual*. Retrieved from http://pebl.sourceforge.net/

Mulckhuyse, M., Talsma, D., & Theeuwes, J. (2007). Grabbing attention without knowing: Automatic capture of attention by subliminal spatial cues. *Visual Cognition*, *15*(7), 779–788.

Müller-Lyer, F. C. (1889). Optische Urteilstäuschungen. *Archiv für Anatomie und Physiologie, Physiologische Abteilung*, *2*, 263–270.

Mullette-Gillman, O., Cohen, Y. E., & Groh, J. M. (2005). Eye-centered, head-centered, and complex coding of visual and auditory targets in the intraparietal sulcus. *Journal of Neurophysiology*, *94*(4), 2331–2352. doi:10.1152/jn.00021.2005

Munoz, D. P., & Wurtz, R. H. (1992). Role of the rostral superior colliculus in active visual fixation and execution of express saccades. *Journal of Neurophysiology*, *67*(4), 1000–1002.

Nakamura, K., & Colby, C. L. (2002). Updating of the visual representation in monkey striate and extrastriate cortex during saccades. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(6), 4026–4031.

Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, *48*(3), 443–453. doi:10.1016/0022-2836(70)90057-4

Noë, A., & O'Regan, J. K. (2000). Perception, attention, and the grand illusion. *Psyche*, *6*(15), 6–15.

Noë, A., Pessoa, L., & Thompson, E. (2000). Beyond the grand illusion: What change blindness really teaches us about vision. *Visual Cognition*, *7*(1-3), 93–106.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. doi:10.1016/j.tics.2006.07.005

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception. *Science*, *171*(3968), 308–311.

Ohtsuka, K. (1994). Properties of memory-guided saccades toward targets flashed during smooth pursuit in human subjects. *Investigative Ophthalmology & Visual Science*, *35*(2), 509–514.

Okuda, T., Tanaka, E., & Kasai, T. (1976). A method for the correction of garbled words based on the Levenshtein metric. *IEEE Transactions on Computers* (Vol. 100, pp. 172–178).

Ong, W. S., & Bisley, J. W. (2011). A lack of anticipatory remapping of retinotopic receptive fields in the middle temporal area. *Journal of Neuroscience*, *31*(29), 10432 –10436. doi:10.1523/JNEUROSCI.5589-10.2011

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.

O'Regan, J. K., Deubel, H., Clark, J. J., & Rensink, R. A. (2000). Picture changes during blinks: Looking without seeing and seeing without looking. *Visual Cognition*, *7*(1), 191–211.

O'Regan, J. K., & Lévy-Schoen, A. (1983). Integrating visual information from successive fixations: does trans-saccadic fusion exist? *Vision Research*, *23*(8), 765–768.

O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, *24*(5), 939–1031.

Overvliet, K. E., Azañon, E., & Soto-Faraco, S. (2011). Somatosensory saccades reveal the timing of tactile spatial

remapping. *Neuropsychologia*, *49*(11), 3046–3052. doi:16/j.neuropsychologia.2011.07.005

Pan, K., & Eriksen, C. W. (1993). Attentional distribution in the visual field during same-different judgments as assessed by response competition. *Perception & Psychophysics*, *53*(2), 134–144.

Parks, N. A., & Corballis, P. M. (2008). Electrophysiological correlates of presaccadic remapping in humans. *Psychophysiology*, *45*(5), 776–783.

Pasley, B. N., David, S. V., Mesgarani, N., Flinker, A., Shamma, S. A., Crone, N. E., Knight, R. T., et al. (2012). Reconstructing speech from human auditory cortex. *PLoS Biology*, *10*(1), e1001251. doi:10.1371/journal.pbio.1001251

Peirce, J. W. (2007). PsychoPy: Psychophysics software in Python. *Journal of Neuroscience Methods*, *162*(1-2), 8–13. doi:10.1016/j.jneumeth.2006.11.017

Pertzov, Y., Zohary, E., & Avidan, G. (2010). Rapid formation of spatiotopic representations as revealed by inhibition of return. *Journal of Neuroscience*, *30*(26), 8882–8887.

Pollatsek, A., Rayner, K., & Collins, W. E. (1984). Integrating pictorial information across eye movements. *Journal of Experimental Psychology: General*, *113*(3), 426–442.

Pollatsek, A., Rayner, K., & Henderson, J. M. (1990). Role of spatial location in integration of pictorial information across saccades. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(1), 199–210.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology*, *32*(1), 3–25.

Posner, M. I., & Cohen, Y. (1984). Components of visual orienting. In H. Bouma & D. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes* (pp. 531–556). Hillsdale, NJ: Erlbaum.

Prime, S. L., Tsotsos, L., Keith, G., & Crawford, J. (2007). Visual memory capacity in transsaccadic integration. *Experimental Brain Research*, *180*(4), 609–628.

Prime, S. L., Vesia, M., & Crawford, J. D. (2011). Cortical mechanisms for trans-saccadic memory and integration of multiple object features. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 540–553. doi:10.1098/rstb.2010.0184

Privitera, C. M., & Stark, L. W. (2000). Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*(9), 970–982.

R Development Core Team. (2010). *R: A language and environment for statistical computing*. Retrieved from http://www.R-project.org/

Rensink, R. A., O'Regan, J. K., & Clark, J. J. (1997). The need for attention to perceive changes in scenes. *Psychological Science*, *8*(5), 368–373.

Reynolds, J. H., & Chelazzi, L. (2004). Attentional modulation of visual processing. *Annual Review of Neuroscience*, *27*, 611–647.

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive mechanisms subserve attention in macaque areas V2 and V4. *Journal of Neuroscience*, *19*(5), 1736–1753.

Reynolds, J. H., Pasternak, T., & Desimone, R. (2000). Attention increases sensitivity of V4 neurons. *Neuron*, *26*(3), 703–714.

Riggs, L. A., Ratliff, F., Cornsweet, J. C., & Cornsweet, T. N. (1953). The disappearance of steadily fixated visual test objects. *Journal of the Optical Society of America*, *43*(6), 495–500.

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

Robbins, N. B. (2004). *Creating More Effective Graphs* (1st ed.). Wiley-Interscience.

Robinson, D. A. (1968). Eye movement control in primates. *Science*, *161*(3847), 1219–1224. doi:10.1126/science.161.3847.1219

Rocha-Miranda, C. E., Bender, D. B., Gross, C. G., & Mishkin, M. (1975). Visual activation of neurons in inferotemporal cortex depends on striate cortex and forebrain commissures. *Journal of Neurophysiology*,

38(3), 475–491.

Rock, I., Linnett, C. M., Grant, P., & Mack, A. (1992). Perception without attention: results of a new method. *Cognitive Psychology*, *24*(4), 502–534.

Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, *395*(6700), 376–381. doi:10.1038/26475

Rolfs, M., Jonikaitis, D., Deubel, H., & Cavanagh, P. (2011). Predictive remapping of attention across eye movements. *Nature Neuroscience*, *14*, 252–256.

Ross, J., Morrone, M. C., Goldberg, M. E., & Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends in Neurosciences*, *24*(2), 113–121.

Rotman, G., Brenner, E., & Smeets, J. B. J. (2005). Flashes are localised as if they were moving with the eyes. *Vision Research*, *45*(3), 355–364. doi:10.1016/j.visres.2004.08.014

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*(7), 631–632.

Sagi, D., & Julesz, B. (1987). Short-range limitation on detection of feature differences. *Spatial Vision*, *2*(1), 39–49.

San Agustin, J., Skovsgaard, H., Mollenbach, E., Barret, M., Tall, M., Hansen, D. W., & Hansen, J. P. (2010). Evaluation of a low-cost open-source gaze tracker. *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 77–80).

Sapir, A., Hayes, A., Henik, A., Danziger, S., & Rafal, R. (2004). Parietal lobe lesions disrupt saccadic remapping of inhibitory location tagging. *Journal of Cognitive Neuroscience*, *16*(4), 503–509.

Sato, T. (1988). Effects of attention and stimulus interaction on visual responses of inferior temporal neurons in macaque. *Journal of Neurophysiology*, *60*(1), 344–364.

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research*, *77*(1), 23–30.

Schlag, J., Schlag-Rey, M., & Dassonville, P. (1990). Saccades can be aimed at the spatial location of targets flashed during pursuit. *Journal of neurophysiology*, *64*(2), 575–581.

Schneider, W. (1988). Micro Experimental Laboratory: An integrated system for IBM PC compatibles. *Behavior Research Methods: Instruments & Computers*, *20*(2), 206–217.

Sereno, A. B., & Kosslyn, S. M. (1991). Discrimination within and between hemifields: a new constraint on theories of attention. *Neuropsychologia*, *29*(7), 659–675.

Shebilske, W. L. (1976). Extraretinal information in corrective saccades and inflow vs outflow theories of visual direction constancy. *Vision Research*, *16*(6), 621–628.

Siebold, A., van Zoest, W., & Donk, M. (2011). Oculomotor evidence for top-down control following the initial saccade. *PLoS ONE*, *6*(9), e23552. doi:10.1371/journal.pone.0023552

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

Smeets, J., Brenner, E., de Grave, D., & Cuijpers, R. (2002). Illusions in action: consequences of inconsistent processing of spatial attributes. *Experimental Brain Research*, *147*(2), 135–144. doi:10.1007/s00221-002-1185-7

Smith, A. T., Singh, K. D., Williams, A. L., & Greenlee, M. W. (2001). Estimating receptive field size from fMRI data in human striate and extrastriate visual cortex. *Cerebral Cortex*, *11*(12), 1182–1190.

Sommer, M. A., & Wurtz, R. H. (2003). The frontal eye field sends predictively remapped visual signals to the superior colliculus. *Journal of Vision*, *3*(9), e146.

Sommer, M. A., & Wurtz, R. H. (2004). What the brain stem tells the frontal cortex. I. Oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *Journal of Neurophysiology*, *91*(3), 1381–1402.

Sommer, M. A., & Wurtz, R. H. (2006). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature*,

444(7117), 374–377.

Sommer, M. A., & Wurtz, R. H. (2008). Brain circuits for the internal monitoring of movements. *Annual Review of Neuroscience*, *31*(1), 317–338.

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience*, *3*(12), 952–964.

Sperry, R. W. (1950). Neural basis of the spontaneous optokinetic response produced by visual inversion. *Journal of Comparative and Physiological Psychology*, *43*(6), 482–489.

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science*, *240*(4850), 338–340.

Srinivasan, M. V., & Bernard, G. D. (1975). The effect of motion on visual acuity of the compound eye: A theoretical analysis. *Vision Research*, *15*(4), 515–525. doi:10.1016/0042-6989(75)90029-2

Stahl, C. (2006). Software for generating psychological experiments. *Experimental Psychology*, *53*(3), 218–232.

Stevens, M., Lammertyn, J., Verbruggen, F., & Vandierendonck, A. (2006). T-scope: A C library for programming cognitive experiments on the MS Windows platform. *Behavior Research Methods*, *38*(2), 280–286.

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096–1104. doi:10.3758/BRM.42.4.1096

Straw, A. D. (2008). Vision Egg: an open-source library for realtime visual stimulus generation. *Frontiers in Neuroinformatics*, *2*, e4. doi:10.3389/neuro.11.004.2008

Talbot, S. A., & Marshall, W. H. (1941). Physiological studies on neural mechanisms of visual localization and discrimination. *American Journal of Ophthalmology*, *24*, 1255–1263.

Tassinari, G., Aglioti, S., Chelazzi, L., Marzi, C. A., & Berlucchi, G. (1987). Distribution in the visual field of the costs of voluntarily allocated attention and of the inhibitory after-effects of covert orienting. *Neuropsychologia*, *25*(1A), 55–71.

Tatler, B. W. (2001). Characterising the visual buffer: Real-world evidence for overwriting early in each fixation. *Perception*, *30*(8), 993–1006.

Tatler, B. W., & Land, M. F. (2011). Vision and the representation of the surroundings in spatial memory. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 596–610. doi:10.1098/rstb.2010.0188

Theeuwes, J. (1991). Exogenous and endogenous control of attention: The effect of visual onsets and offsets. *Perception & Psychophysics*, *49*(1), 83–90.

Theeuwes, J. (1994). Stimulus-driven capture and attentional set: Selective search for color and visual abrupt onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 799–799.

Theeuwes, J. (2010). Top-down and bottom-up control of visual selection. *Acta Psychologica*, *135*, 77–99.

Theeuwes, J., Atchley, P., & Kramer, A. F. (2000). On the time course of top-down and bottom-up control of visual attention. In S. Monsell & J. Driver (Eds.), *Attention and Performance XVIII: Control of Cognitive Processes* (pp. 105–124). Cambridge, MA: MIT Press.

Theeuwes, J., Kramer, A. F., Hahn, S., & Irwin, D. E. (1998). Our eyes do not always go where we want them to go: Capture of the eyes by new objects. *Psychological Science*, *9*(5), 379–385.

Theeuwes, J., Mathôt, S., & Grainger, J. (in preparation). *Object-based Cueing Effects*.

Theeuwes, J., Mathôt, S., & Kingstone, A. (2010). Object-based eye movements: The eyes prefer to stay within the same object. *Perception & Psychophysics*, *2009*, 12–21.

The MathWorks. (1998). *MATLAB Use Guide*.

Thier, P., Haarmeier, T., Chakraborty, S., Lindner, A., & Tikhonov, A. (2001). Cortical substrates of perceptual stability during eye movements. *NeuroImage*, *14*(1), 33–39. doi:10.1006/nimg.2001.0840

Tiobe.com. (2011, May). Most Popular Programming Languages. Retrieved from http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html

Tipper, S. P., Howard, L. A., & Jackson, S. R. (1997). Selective reaching to grasp: Evidence for distractor interference

effects. *Visual Cognition*, *4*(1), 1–38.

Tolias, A. S., Moore, T., Smirnakis, S. M., Tehovnik, E. J., Siapas, A. G., & Schiller, P. H. (2001). Eye movements modulate visual receptive fields of V4 neurons. *Neuron*, *29*(3), 757–767.

Tombu, M., & Tsotsos, J. K. (2008). Attending to orientation results in an inhibitory surround in orientation space. *Perception & Psychophysics*, *70*(1), 30–35.

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, *21*, 753–759.

Torralbo, A., & Beck, D. M. (2008). Perceptual load-induced selection as a result of local competitive interactions in visual cortex. *Psychological Science*, *19*(10), 1045–1050.

Trehub, A. (1977). Neuronal models for cognitive processes: networks for learning, perception and imagination. *Journal of Theoretical Biology*, *65*(1), 141–169.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*(1), 97–136.

Tsotsos, J. K., Culhane, S. M., Kei Wai, W. Y., Lai, Y., Davis, N., & Nuflo, F. (1995). Modeling visual attention via selective tuning. *Artificial Intelligence*, *78*(1-2), 507–545.

Ulrich, R., & Giray, M. (1989). Time resolution of clocks: Effects on reaction time measurement—Good news for bad clocks. *British Journal of Mathematical and Statistical Psychology*, *42*(1), 1–12. doi:10.1111/j.2044-8317.1989.tb01111.x

Umeno, M. M., & Goldberg, M. E. (1997). Spatial processing in the monkey frontal eye field. I. Predictive visual responses. *Journal of Neurophysiology*, *78*(3), 1373–1383.

Umiltà, C., Riggio, L., Dascola, I., & Rizzolatti, G. (1991). Differential effects of central and peripheral cues on the reorienting of spatial attention. *European Journal of Cognitive Psychology*, *3*(2), 247–267.

Vainio, L., Ellis, R., & Tucker, M. (2007). The role of visual attention in action priming. *Quarterly Journal of Experimental Psychology*, *60*(2), 241. doi:10.1080/17470210600625149

van Koningsbruggen, M. G., Gabay, S., Sapir, A., Henik, A., & Rafal, R. D. (2010). Hemispheric asymmetry in the remapping and maintenance of visual saliency maps: a TMS study. *Journal of Cognitive Neuroscience*, *22*(8), 1730–1738. doi:10.1162/jocn.2009.21356

Van Rossum, G. (2008). *Python reference manual*. Python software foundation.

Volkmann, F. C., Riggs, L. A., White, K. D., & Moore, R. K. (1978). Contrast sensitivity during saccadic eye movements. *Vision Research*, *18*(9), 1193–1199.

von Holst, E., & Mittelstaedt, H. (1950). Das Reafferenzprinzip. *Naturwissenschaften*, *37*(20), 464–476.

Wagner, R. A., & Lowrance, R. (1975). An extension of the string-to-string correction problem. *Journal of the ACM*, *22*(2), 177–183.

Walker, M. F., Fitzgibbon, E. J., & Goldberg, M. E. (1995). Neurons in the monkey superior colliculus predict the visual result of impending saccadic eye movements. *Journal of Neurophysiology*, *73*(5), 1988–2003.

Walls, G. L. (1962). The evolutionary history of eye movements. *Vision Research*, *2*(1-4), 69–80.

Wang, Z., Satel, J., Trappenberg, T. P., & Klein, R. M. (2011). Aftereffects of saccades explored in a dynamic neural field model of the superior colliculus. *Journal of Eye Movement Research*, *4*(2), 1–16.

Wannig, A., Stanisor, L., & Roelfsema, P. R. (2011). Automatic spread of attentional response modulation along Gestalt criteria in primary visual cortex. *Nature Neuroscience, 14*, 1243–1244. doi:10.1038/nn.2910

Wenderoth, P., & Wiese, M. (2008). Retinotopic encoding of the direction aftereffect. *Vision Research*, *48*(19), 1949–1954.

West, J. M., Haake, A. R., Rozanski, E. P., & Karn, K. S. (2006). eyePatterns: software for identifying patterns and similarities across fixation sequences. *Proceedings of the 2006 Symposium on Eye Tracking Research & Applications* (pp. 149–154). San Diego, California: ACM. doi:10.1145/1117309.1117360

Wetzels, R., Matzke, D., Lee, M. D., Rouder, J. N., Iverson, G. J., & Wagenmakers, E.-J. (2011). Statistical evidence in experimental psychology. *Perspectives on Psychological Science*, *6*(3), 291–298.

doi:10.1177/1745691611406923

Wittenberg, M., Bremmer, F., & Wachtler, T. (2008). Perceptual evidence for saccadic updating of color stimuli. *Journal of Vision*, *8*(14), e9.

Wurtz, R. H. (2008). Neuronal mechanisms of visual stability. *Vision Research*, *48*(20), 2070–2089.

Yantis, S., & Hillstrom, A. P. (1994). Stimulus-driven attentional capture: evidence from equiluminant visual objects. *Journal of Experimental Psychology: Human Perception and Performance*, *20*(1), 95–107.

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 601–621.

Yantis, S., & Jonides, J. (1996). Attentional capture by abrupt onsets: New perceptual objects or visual masking? *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1505–1513.

Zangemeister, W. H., & Liman, T. (2007). Foveal versus parafoveal scanpaths of visual imagery in virtual hemianopic subjects. *Computers in Biology and Medicine*, *37*(7), 975–982. doi:10.1016/j.compbiomed.2007.01.015

Zangemeister, W. H., & Oechsner, U. (1996). Evidence for scanpaths in hemianopic patients shown through string editing methods. *Advances in Psychology*, *116*, 197–221.

Zimmermann, E., Burr, D., & Morrone, M. C. (2011). Spatiotopic visual maps revealed by saccadic adaptation in humans. *Current Biology*, *21*(16), 1380–1384. doi:16/j.cub.2011.06.014

Zimmer, M., & Kovács, G. (2011). Position specificity of adaptation-related face aftereffects. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 586 –595. doi:10.1098/rstb.2010.0265

Zirnsak, M., Gerhards, R. G. K., Kiani, R., Lappe, M., & Hamker, F. H. (2011). Anticipatory saccade target processing and the presaccadic transfer of visual features. *Journal of Neuroscience*, *31*(49), 17887–17891. doi:10.1523/JNEUROSCI.2465-11.2011

Zirnsak, M., Lappe, M., & Hamker, F. H. (2010). The spatial distribution of receptive field changes in a model of peri-saccadic perception: Predictive remapping and shifts towards the saccade target. *Vision Research*, *50*(14), 1328–1337. doi:10.1016/j.visres.2010.02.002

XVI. Acknowledgements

First off, I would like to thank Jan, who has been a great supervisor. He has always been quick to help, comment, offer ideas for experiments, and skilfully guide our papers through the minefield of academic publishing. But he has never been quick to interfere or criticise, even when my projects digressed substantially from the original plan. This thesis marks the end of Jan as my PhD supervisor, but not, I hope, as my close collaborator and friend.

I would like to thank the rest of my colleagues in the order in which I met them. As a student I never really considered pursuing a career in academia, until I met Jan Berend, who supervised my Bachelor's project in his own quirky way. For some reason, despite a mediocre thesis[19], he suggested that I do a research master, instead of the more conventional clinically oriented one. Because of his suggestion, I approached Mieke, who was at that time setting up the Research Master Cognitive Neuropsychology. I applied, but unfortunately I was almost the only one, and the start of the master was postponed. I think it characterises Mieke (and the department as a whole) that she went out of her way to set up an ad hoc curriculum, so that I nevertheless graduated from a non-existent research master two years later. During these two years, I attended inspiring lectures by Artem, Martijn, Mark, Dirk, and Chris (among others), and I remember with particular fondness the courses Behavioural Methods, by Sander, and Seminar Cognitive Neuroscience, by Durk. The papers that I wrote for these courses made being student feel a bit like being a scientist.

Mieke suggested that I go abroad for my final master's project, which I did. After consulting Stefan, who I thought was the youngest looking professor ever but nevertheless addressed with due courtesy, I went to Alan Kingstone's lab in Vancouver to do a project with Jan, who was there on a sabbatical (but always working!). The BAR lab felt just like home, because of Alan's hospitality (of course), but also because it literally was like home. At some point most lab members were Dutch: Jan, Wieske, Clayton (well, partly), Lotje, and myself.

---

19 I came up with a weird design to investigate the relationship between intelligence and inhibitory skill. Nothing came out of it, except that women are generally better than men (trivial), and that my friends are not nearly as intelligent as they believed.

∞

XVII. Author Publications

Mathôt, S., & Theeuwes, J. (in press). A reinvestigation of the reference frame of the tilt-adaptation aftereffect. *Scientific Reports*.

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*, 314-324.

Mathôt, S., & Theeuwes, J. (2012). It's all about the transient: Intra-saccadic onset stimuli do not capture attention. *Journal of Eye Movement Research*, *5*(2):4, 1-12.

Mathôt, S., Cristino, F., Gilchrist, I.D., & Theeuwes, J. (2012). A simple way to estimate similarity between pairs of eye movement sequences. *Journal of Eye Movement Research*, *5*(1):4, 1-15.

Mathôt, S., & Theeuwes, J. (2011). Mantra: An open method for object and movement tracking. *Behavior Research Methods*, *43*(4), 1182-1193.

Mathôt, S., & Theeuwes, J. (2011). Visual attention and stability. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *366*(1564), 516-527.

Cristino, F., **Mathôt, S.**, Theeuwes, J., & Gilchrist, I. D. (2010). ScanMatch: A novel method for comparing fixation sequences. *Behavior Research Methods*, *42*(3), 692-700.

Mathôt, S., Hickey, C., & Theeuwes, J. (2010). From reorienting of attention to biased competition: Evidence from hemifield effects. *Attention, Perception & Psychophysics*, *72*(3), 651-657.

Mathôt, S., & Theeuwes, J. (2010). Gradual remapping results in early retinotopic and late spatiotopic inhibition of return. *Psychological Science*, *21*(12), 1793-1798.

Mathôt, S., & Theeuwes, J. (2010). Evidence for the predictive remapping of visual attention. *Experimental Brain Research*, *200*(1), 117-122.

Theeuwes, J., **Mathôt, S.**, & Kingstone, A. (2010). Object-based eye movements: The eyes prefer to stay within the same object. *Attention, Perception & Psychophysics*, *72*(3), 12–21.