Internal representations of the brain : shortterm visual memory and tool integration

Internal Representations of the Brain

Short-term visual memory and tool integration

Rosanne L. Rademaker

ISBN 978–90–9029462–9

Internal Representations of the Brain

Short-term visual memory and tool integration

DISSERTATION

to obtain the degree of Doctor at Maastricht University, on the authority of the Rector Magnificus, Prof. Dr. L.L.G. Soete in accordance with the decision of the Board of Deans, to be defended in public on Thursday the 10th of December 2015 at 16.00 hrs.

by

Rosanne L. Rademaker

Promotors
Prof. Dr. Alexander T. Sack

Prof. Dr. Peter De Weerd

Copromotor

Dr. Janneke F.M. Jehee

Assessment Committee

Prof. Dr. Bernadette M. Jansma (Chair)

Prof. Dr. Rainer W. Goebel

Dr. Sam Ling (Boston University)

Prof. Dr. Pieter R. Roelfsema (Netherlands Institute for Neuroscience)

The work in this thesis was supported by grants from the Netherlands Organization of Scientific Research (NWO) as well as the European Research Council (ERC)

Contents

Chapter 1

General Introduction

Introduction

The ultimate purpose of the human brain could be formulated as an optimization process – processing sensory input in order to produce behaviors maximally beneficial to the organism. Viewed as such, the brain is a simple input-output machine designed to connect environmental circumstances gauged via the senses with the obtainment of behavioral goals via motor output. William James (1890), in reference to the function of the hemispheres captured this idea by noting that: "The highest centres do probably contain nothing but arrangements for representing impressions and movements, and other arrangements for coupling the activity of these arrangements together". Effectively, James (1890) described the hemispheres as nothing more than a complex relay between input (impressions) and output (movements), the idea being that such an arrangement protracts, mediates and complicates the reflexive actions of lower centers by adding some extra steps in between (James, 1890).

Something indeed happens in between, and the architecture of the mammalian cortex is tantamount to the great lengths the evolutionary process has gone through to insert these 'extra steps'. For example, a single spiny stellate cell in cat visual cortex has about 5000 synapses, but only 5-6% of connections come from thalamus, while the rest comes from other cortical cells (Da Costa & Martin, 2009). Overall, less than 1% of excitatory connections are coming from an external input (Kennedy, Knoblauch, & Toroczkai, 2013; Schüz & Braitenberg, 2002). And if that is not bad enough, consider this: 1 cubic mm of mouse white matter contains enough axonal fibers to make a string spanning a distance of 9 meter, but 1 cubic mm of grey matter provides enough axonal fixers to make a string of 4.1 km! (Braitenberg & Schüz, 1991). Sensory input is of course important, and pathways with small numbers of fibers may have a big input because their synchronous and correlated firing is not easily ignored (Da Costa & Martin, 2011). However, the message here is that there is a massive recurrence within cortex and the vast majority of connections exist in local cortical circuits. Most of what cortex is doing is really just about stuff it's making up on its own, without a direct link to input or output…

"Every sensorial excitement propagated to a lower centre tends to spread upwards and arouse an idea – every idea tends ultimately either to produce a movement or to check one which otherwise would be produced". Here, James (1890) metaphysically refers to the 'extra steps' as 'ideas'. Alternatively, one can refer to them as thoughts, cognition, or internal representations. One may wonder at all the computational processing happening in that 1 cubic mm of grey matter, or within any of the other cortical implementations of the 'extra steps' – and be in awe. What *is* clear is that, somehow, this profound computational core drives what makes us human; it allows us to make important life choices, such as where to live, which partner to choose, or what to dedicate our professional life to; it allows us to feel overwhelming emotions, like paralyzing fear, tearful laughter, or intense sadness.

While most computations might occur in the absence of direct sensory input, many of our cognitive experiences have a sensory vibe to them nonetheless. Take a second to consider that moment during which you suddenly remembered the face of a long lost friend, or the scent of your childhood shampoo. While lacking a direct connection to the present world, such internal sensory reproductions can appear strikingly lifelike. Internal sensory experiences can also be invoked at will, to fulfill a specific task. In search of your keys it might help to create a vivid mental image of the last place you saw them. Trying to cross a busy road you want to hold the various speeds and directions of many cars in mind all at once to ensure a safe passage while watching where you're going.

Broadly speaking, the main focus of the current thesis is to investigate the 'extra steps' that allow humans to keep visual representations online in the absence of direct sensory input. Many everyday tasks like the ones described above critically depend on a veridical visual buffer, allowing people to briefly maintain visual information for some future use. While seemingly simple, holding on to even basic visual features already poses some serious computational challenges. Specifically, we will focus on mechanisms of storage over time, and the ability to maintain robust sensory representations despite the constant stream of

potentially conflicting information arriving at the eyes. How might visual memories be inoculated against interference and temporal decay? Moreover, how are the 'extra steps' required for the maintenance of high fidelity sensorial representations implemented in cortex?

The second goal of this thesis veers more closely to the output side of things, and is about investigating the representation of the human body and the representation of tools. Imagine you managed to find your keys and now you need to lock the door. Upon grabbing your keys you have an immediate sense of where they are in space and in relation to your body, and presumably a swift and automated hand gesture is all it takes to insert the key and turn the lock. For humans in particular many behavioral goals require the use of tools. Therefore, the ability to flexibly update the body's representation to also represent a tool is of pivotal importance to perform the motor acts required to achieve such goals. In contrast with visual memory, the question of tool integration has input and output lying much closer together, with sensations and movements highly interwoven. Similar to visual memory, various 'extra steps' are required of cortex, as a representation must be formed of the tool even though no sensors are present on it to supply direct sensorial context. How flexible are representations of extensions that are not the body's own, and how rapidly might they emerge?

Throughout this general introduction I will provide some relevant background to understanding brain processes involved in representations of information in the absence of direct physical input. Such representations can be extended in time, as with the online maintenance of visual information after it is no longer directly accessible to the eyes. Such representations can also be extended in space, as when trying to modify the body representation to incorporate external objects. First, we will discuss visual working memory, broadly outlined conform Marr's three complementary levels of analysis in information processing systems (Marr, 1982). We will explore (1) the computations required for the online maintenance of visual information, discussing what the system does and the problems it needs to overcome. Such problems include, for example, the

severe limitations of the system in terms of quantity. Then we turn to (2) representations and processes involved in memory maintenance, and how such representations might be probed. A closer look at (3) the implementation of a visual memory system in dispersed cortical networks will be provided next, focusing on the recruitment of visual sensory areas in particular.

Finally, we will briefly consider ways in which the work presented in this thesis advances our understanding of specific top-down cognitive states: the short-term maintenance of visual representations, and the inferences required to incorporate tools as part of the body representation.

Challenges of a visual working memory system

"Seeing is believing", as the saying goes, implies that our visual world is an accurate reflection of reality. A reality that we can believe and trust once perceived with our own eyes. Strictly speaking however, visual acuity is high only within the central most region of visual space (Gruesser & Gruesser-Cornehls, 1978), and eye movements are made roughly every 300 milliseconds (Rayner, 1978). Thus, the information sensed with our eyes really consists of only a small region of space that jitters around at a fairly high frequency. Luckily, the brain is a powerful system permitting a stable, continuous, and rich percept despite the discontinuous and sparse nature of the information gauged by its sensors. When reconstructing a truthful visual world, one factor of critical importance is a temporal buffer that allows visual information to be stored in the absence of physical input. Besides its potential for reconstructing our visual world on the fly, the visual buffer is often investigated in the context of aiding the attainment of cognitive goals.

The visual buffer is often referred to as visual 'short-term' or 'working' memory. In contrast to the high capacity, but fragile and fleeting iconic memory store (Sperling, 1960), or a recently proposed intermediate visual store of relatively high capacity (Sligte, Scholte, & Lamme, 2008; 2009), visual short-term memory is believed to represent a limited

amount of information over extended periods of time in a highly robust manner (Phillips, 1974). The terms 'short-term' and 'working' memory are often used interchangeably, though traditionally they describe different segments of the literature. The former is a more theory-neutral way to describe information maintenance over time; the latter originated from a theoretical system composed of a central executive and several sensory 'slave systems' for temporary information storage (Baddeley & Hitch, 1974). While we should recognize the subtle differences between these two terms, here they will be used more or less synonymously to refer to the *online maintenance of visual information* under varying task circumstances – it can be argued that even the simplest short-term memory tasks require some degree of information monitoring in order to fulfill task requirements, illustrating that the line between 'short-term' and 'working' memory is a blurry one.

Besides its severe *capacity limitations*, often estimated around 3-4 simple visual items (Luck & Vogel, 1997; Zhang & Luck, 2008) which we shall discuss in more detail later, it is pivotal for the working memory system to store representations *veridically* if it is to adequately realize its perceptual and cognitive goals. The importance of visual working memory to higher cognition and everyday functioning is exemplified by work showing that the ability to store high-resolution visual representations is protracted over the course of development, with increased precision as children become older (Burnett Heyes, Zokaei, Van Der Staaij, Bays, & Husain 2012). Furthermore, memory performance is correlated with general cognitive ability (Johnson et al., 2013), as well as fluid intelligence (Kane & Engle, 2002) – larger memory capacity being associated with higher fluid intelligence (Fukuda, Vogel, Mayr, & Awh, 2010). The ability to suppress irrelevant inputs seems to mediate this relationship between fluid intelligence and working memory capacity (Burgess, Gray, Conway, & Braver, 2011), signifying a third major challenge to the system, namely the *inoculation of memories against interference*. Finally, *persistent representations* must be achieved over time even when there is no ground-truth percept to anchor these representations.

Capacity limitations. A key avenue of investigation into visual working memory

system has centered on the system's profound quantity limitations: why does performance suffer so drastically when more items are committed to memory? Often, a discrete upper quantity is invoked to signify the number of items that one can faithfully maintain in memory. This number, as mentioned before, is generally estimated at 3-4 items (Cowan, 2000; Luck & Vogel, 1997; Vogel, Woodman, & Luck, 2001; Zhang & Luck, 2008). However, such a conceptualization of capacity is in no way unique to memory, and comparable limits have been demonstrated in various other domains of psychophysical inquiry. Examples include multiple object tracking (Pylyshyn & Storm, 1988) and visual search (Alvarez & Cavanagh, 2004), the effect of subitizing in numerosity (Kaufman, Lord, Reese, & Volkmann 1949; Trick & Pylyshyn, 1994), as well as processing limits resulting in change blindness (Simons & Levin, 1997).

Capacity limits in working memory might thus reflect a more central bottleneck of information processing, shared by many cognitive activities. The real jam might ensue at the perceptual / attentional end of things, simply getting carried over from encoding into memory (Buschman, Siegel, Roy, & Miller, 2011). Indeed, it's been suggested that once items are committed to memory the delay between a sample and a test in a visual memory task barely alters performance, with longer delays resulting in negligible loss of precision (Magnussen, Greenlee, Asplund & Dyrnes, 1991; Magnussen & Greenlee, 1992; Regan, 1985). However, the extent to which memory capacity is being taxed does interact with the duration of the memory delay, with faster decay when a larger number of items (Pearson, Raškevičius, Bays, Pertzov, & Husain, 2014) or more complex items (Phillips, 1974) need to be remembered.

Recently there has been a surge of attempts to explain and model the capacity limitations of visual working memory. We will briefly review some influential models while keeping the aforementioned idea of a general bottleneck in mind. In fact, the generality of capacity limitations might even bolster the implications of visual working memory models, owing to their potential to be applied to other bandwidth-limited cognitive processes in the future. A typical working memory task briefly presents participants with an array of one

or several items that have to be retained for several seconds. Items are often independently drawn samples from some continuous variable such as color or orientation. In change detection paradigms, a second display will appear after the delay and participants indicate whether or not one of the items changed relative to the first display. In method of adjustment paradigms, the second display generally consists of a location cue indicating which of the items will need to be reconstructed by adjusting a test stimulus (Wilken & Ma, 2004). The models discussed next have primarily employed the method of adjustment procedure, and its resultant error distributions of report, to see how well behavior can be explained.

First of all, a broad distinction can be made between deterministic and stochastic models of capacity limits (Bays, 2015). The first class assumes a discrete limit on the number of items that can be remembered. A classic example is the infamous 'slot model' already alluded to above, which suggests that items can be stored in 3-4 discrete slots, each of which has a fixed resolution (Luck & Vogel, 1997). This model predicts that memory precision is stable irrespective of the number of items in memory, but once the number of items exceeds the number of available slots the excess items are invariably forgotten. An extension of this model is the 'slots and averaging' model (Zhang & Luck, 2008). As with the classic slot model, multiple independent representations are assumed. Additionally, when the number of memory items is smaller than the number of slots, several slots can be recruited to hold the same representation. Because each representation has a normally distributed error, averaging together duplicate representations for a single item allocated several slots will result in a lower variability for that item at recall. Thus, while the slots and averaging model still maintains a deterministic upper limit on the number of items that can be remembered, it also assumes that resources can be shared in a quantized manner when there are fewer items than slots.

Instead, stochastic models adhere to the premise that representations become increasingly noisy as more of them are remembered. Contrary to the predictions of deterministic models (i.e. stable precision together with forgetting), stochastic models assume changes

in variability but not guesses – although with very many items in memory, representations might come to resemble noise. This idea is exemplified by the 'resource model', which assumes a shared pool of resources that can be distributed amongst as many items as needed (Bays & Husain, 2008; Bays, Catalao, & Husain, 2009). A variant on this model is the 'variable precision model', which assumes that resources can be flexibly distributed from trial-to-trial and from item-to-item (van den Berg, Shin, Chou, George, & Ma, 2012; Fougnie, Suchow, & Alvarez, 2012). In this model, each item's representation can be considered distributed along a von Mises distribution (the circular equivalent of a normal distribution), and several items can be represented as several overlapping von Mises distributions, each with a variable degree of precision. By considering a representation comprised of many overlapping von Mises functions, observations typical to working memory data can be explained, such as the signature 'peakedness' of response distributions (van den Berg et al., 2012; Fougnie et al., 2012) signified by a kurtosis that is higher than that of your typical von Mises distribution. The variable precision model explains set size effects as an increased variability amongst items as more of them have to be remembered.

More recently, a 'population coding model' has been suggested based on stochastic neuronal firing in a population of neurons, combined with a global normalization process (Bays, 2014). The normalization component of this model explains set size effects by assuming a constant firing rate across the neuronal population irrespective of the number of items in memory. Here, more representations mean fewer spikes per representation, making it harder to decode individual items from the population response, in turn resulting in a variability increase. Normalization is classically thought of as a canonical neural computation that occurs quite locally – by dividing a neuron's response with the summed activity of a pool of neurons with similar receptive fields or tuning properties (Carandini & Heeger, 2011; Heeger, 1992). Nevertheless, there is tentative evidence to support the idea of global normalization. For example, activity in neuronal populations became less informative about the contents of memory when more items were remembered (Buschman, Siegel, Roy, & Miller, 2011; Sprague, Ester, & Serences, 2014).

Besides providing a plausible biological basis for why memory resources are limited, population coding models can be used to explain persistent firing over time (Goldman-Rakic, 1995). We will discuss population coding models and their relation to persistent firing in more detail later on.

Representations of a visual working memory system

Here we will explore some potential representational schemes employed by the visual memory system in order to meet its everyday requirements. We will also explore how one might go about investigating such representations. In fact, by discussing the various models of memory quantity above we have already unveiled some possible systems of representation, such as the notion that circular variables are represented by von Mises distributions. We have even hinted at the possible implementation of a memory system in a neural population code. Here we will take these ideas beyond the question of capacity limitations, and also consider situations in which capacity is not exhausted.

Even with only a single simple visual item in memory there will be a certain degree of internal variability – no system is ever perfect, and neither are most memory representations. We will consider memory quality as the amount of variability around the correct representation, with less variability meaning a higher quality representation. While it is easy to see how this would hold true for continuous variables, it is somewhat less obvious for categorical variables. Consider a memory of whether something was upstairs or downstairs, a mouse or an elephant, a yes or a no – such variables allow little interpolation. Therefore our focus is on representations of simple and continuous visual features, which can be presented with a certain degree of variability in a meaningful way.

Interestingly, some recent work has managed to probe internal variability by simply asking people about it. In doing so, Rademaker et al. (2012) showed that memory representations of visual orientation were variable from trial-to-trial, but also that people had conscious access to the variability of their internal representations (Rademaker,

Tredway, & Tong, 2012). In this study participants were briefly presented with 3 or 6 target orientations, of which one was probed for report after a 3s delay. A test grating was rotated by means of button presses to match the probed orientation. In addition, participants were asked about their confidence on a 6-point scale. Key was that participants indicated their uncertainty on very many trials of the same set size, and varying degrees of confidence were reported from one trial to the next, regardless. These self-reports predicted the quality of visual representation, with memory representations being more variable at higher levels of uncertainty (Rademaker et al., 2012).

The finding that internal representations are variable, and that people have conscious access, is not unique to visual working memory. It's been shown that metacognition predicts variability of single instances of visual mental imagery strength as well (Pearson, Rademaker, & Tong, 2011). Moreover, such metacognitive access can even improve with training (Rademaker & Pearson, 2012). Thus, metacognitive judgments provide a neat way of assessing representational uncertainty, and the quality of internal representations. Besides, such work can inform quantity related inquiries as well. For example, only a model incorporating variable precision, but not slot or fixed-precision models, can predict the kind of trial-to-trial variability within a single set size demonstrated by Rademaker et al. (2012).

Sustained activity. One of the most important requirements of the visual memory system is to maintain representations in the absence of direct sensory input. How are veridical internal representations maintained when there is no objective reality against which to test the validity of such representations? Furthermore, how might such representations change as a function of time, are they slowly corrupted with noise (Kinchla & Smyzer, 1967; Lee & Harris, 1996), or might they terminate suddenly in an all-or-none manner (Zhang & Luck, 2009)? In Chapter 4 the latter question will be addressed empirically, so for now we will primarily discuss the former question in the broader context of how persistent activity could be achieved at a systems level, already hinting at the implementation within the brain's circuitry. We will briefly discuss three options: The

first concerns sustained firing, the second revisits population coding models, and the third challenges the idea of neuronal spiking as a basis for memory maintenance altogether.

The first idea is rather simple, and assumes that those neurons that fire selectively for a perceived item maintain their selective activity in the absence of direct visual input. Evidence for persistent firing to mnemonic items comes primarily from neurons residing in lateral prefrontal cortex. For example, while rhesus macaques were waiting to saccade to a remembered location, neuronal responses in the dorsolateral prefrontal cortex of the animals showed sustained neuronal activity coding for the memorized location (Funahashi, Bruce, & Goldman-Rakic, 1989). Similarly, sustained firing of neurons in macaque prefrontal cortex was found to selectively reflect memories of complex real world objects (Miller, Li, & Desimone, 1993; Miller, Erickson, & Desimone, 1996). Other primate work has additionally implicated temporal cortex as a site where sustained firing can be found (Ranganath & D'Esposito, 2005). For example, while monkeys were remembering a color, neurons recorded from the lower bank of the superior temporal sulcus demonstrated color dependent sustained firing during the delay (Fuster & Jervey, 1981; 1982). However, persistent activity typically disappears once a second task is performed during the delay interval (Miller, Li, & Desimone, 1993; Watanabe & Funahashi, 2014), begging the question if persistent firing really is the central mechanism by which information is maintained over time.

A second option explores the possibility of population coding as a mechanism by which information is maintained over time (Sreenivasan, Curtis, & D'Esposito, 2014; Stokes, 2015). Population coding models, or 'neural network models', can be static or dynamic in nature, and are considered neurally plausible. These models exploit recurrent connections between neurons in distributed neural populations, generally connected through locally excitatory and laterally inhibitory connections. Activity can be sustained over time when connections between neurons form a so-called 'attractor state', which is a stable pattern of firing supported by the network's connectivity. In dynamic neural network models this pattern is represented dynamically, meaning that the pattern of population activity

undergoes both spatial and temporal changes. In other words, the 'state' of a population of neurons that encode mnemonic information at a given time point might be entirely different from the state of the network at second time point, and follows a dynamic trajectory. Ultimately, both static and dynamic neural network models infer that representations are maintained through the state of neuronal activity within the network.

There is accumulating evidence to support the notion that mnemonic information is maintained in a temporally dynamic population code. For example, decoding mnemonic information from neuronal activity in prefrontal populations worked best when performed on activity from concurrent or adjacent points in time, while decoding performance dropped off rapidly with activity from more distant time points (Meyers, Freedman, Kreiman, Miller, & Poggio, 2008; Stokes, Kusunoki, Sigala, Nili, Gaffan, & Duncan, 2013). Thus, information represented in the network's activation state was dynamically changing over time. Notably, besides supporting the online maintenance of information, population codes might more generally serve as a canonical mechanism supporting functionality all over the brain. On the sensory end of things, information about a perceived orientation was reliably decoded from the combined activity of a population of orientation selective neurons in macaque primary visual cortex (Graf, Kohn, Jazayeri, & Movshon, 2011). At the other end of the spectrum, detailed information about a pending movement was similarly stored in a population code, and the combined action of an entire population of broadly tuned motor neurons could predict detailed reaching movements (Georgopoulos, Schwartz, & Kettner, 1986).

Finally, it's been suggested that spiking may not be the sole mechanism responsible for carrying information across temporal intervals (Mongillo, Barak, & Tsodyks, 2008). Instead, it was theorized that in a network of excitatory connections, neurons coding for a mnemonic item could change their synaptic weights through elevated presynaptic calcium levels. Such transient synaptic facilitation can presumably be maintained for about 1 second without any spiking, and read out by a sweep of spiking activity through the network (Mongillo et al., 2008). Such an account could explain reports of transient, but

not sustained, neuronal spiking during a memory delay interval (Barak, Tsodyks, & Romo, 2010; Shafi, et al., 2007).

Implementation of a visual working memory system

"The same cerebral process which, when aroused from without by a sense-organ, gives the perception of an object, will give an idea of the same object when aroused by other cerebral processes from within" – James, 1890

Sensory recruitment. At the advent of modern day psychology an idea was coined, probably not for the first time even then, about the involvement of the brain's sensory centers in top down cognitive processes. This idea has been rehashed, reformulated, and reinvented many times over; many areas of inquiry have concerned themselves with it; and while fleshing out the details for one cognitive function or another, the idea has kept its close resemblance to earlier and parallel accounts. At the core, the idea is that activity in sensory areas is reinstated during top-down cognitive processes that contain elements relating to the original sensory experience. This idea of sensory recruitment is intuitively appealing – it makes sense for higher-level executive areas to recruit sensory areas, specialized in processing the sensory analogs of cognitive contents.

In memory research this idea entails a movement away from supposing specialized 'memory units', and towards the notion that the coordinated effort of executive and sensory components is central to mnemonic function. This has been formalized in various ways over the years. For example, the influential working memory model from cognitive psychology developed by Baddeley and Hitch (1974) postulates a supervisory 'central executive' supported by sensory-like 'slave systems' responsible for the short-term storage of phonological and visuo-spatial information. Another of many possible examples is Crowder (1993) who wrote on the topic of auditory memory: "Where that original experience was played out, in brain activity, is where the memory for it will correspondingly reside afterwards", echoing the words of William James more than a

century earlier. Also in the field of long-term memory the idea pops up, with 'cortical reinstatement theory' positing that memory retrieval is accomplished when lower-level sensory areas, that also represented the initial sensory counterparts of mnemonic content, are reactivated (Nyberg, Habib, McIntosh, & Tulving, 2000; Wheeler, Petersen, & Buckner, 2000). All iterations on sensory recruitment imply a sensory specific implementation of the memory system in the brain – if a memory was visually encoded, visual sensory areas will be activated to represent it during retrieval, while auditory memories will reactivate auditory cortex during retrieval, and so on.

Sensory recruitment theory provides a tentative answer to how the brain as a whole meets the computational demands associated with the maintenance of information to which it no longer has access. In its crudest form, the answer is that it requires well-orchestrated interactions between multiple areas of the brain, each playing their own part. From our previous deliberations on sustained mnemonic representations over time we gleaned what single neurons in anterior parts of the brain, as well as local recurrent networks, might bring to the table. We will return to the actions of the entire network during working memory at the end of this section. But first, we will discuss the contribution of early sensory regions, and early visual cortex in particular. We will also briefly touch upon another top-down cognitive state, feature-based attention, and how we attempted to explore its effects on early visual processing. Then, we will invert our perspective, switching from the effects of memory (or attentional) contents on visual processing, to the effects of visual processing on memory representations. After all, the inoculation of visual memories against interference from the eyes is one of the major challenges of the visual working memory system.

Early visual cortex. Recent neuroimaging has provided ample evidence showing that areas involved in sensory processing of a certain stimulus are also involved when processing information about that stimulus in its absence, without direct sensory input. For example, both visual mental imagery (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Kosslyn et al., 1999; Lee, Kravitz, & Baker, 2012; O'Craven & Kanwisher, 2000) and

the visual contents of dreams (Horikawa, Tamaki, Miyawaki, & Kamitani, 2013) rely on information processing by the visual sensory system, as do long-term visual memories (Bosch, Jehee, Fernández, & Doeller, 2014). Such sensory recruitment holds true for other modalities as well, where internally generated tactile or auditory experiences involve early somatosensory and auditory areas respectively (Harris, Miniussi, Harris, & Diamond, 2002; Meyer et al., 2010). The contents of working memory for simple visual features such as visual orientation (Ester, Serences & Awh, 2009; Harrison & Tong, 2009), color (Serences, Ester, Vogel, & Awh, 2009), motion (Riggal & Postle, 2012), or location (Sprague, Ester, & Serences 2014) rely on representations in primary visual cortex. For example, patterns of activity in visual sensory cortex, including primary visual area V1, were predictive of the orientation someone was remembering (Harrison & Tong, 2009; Serences et al., 2009). Moreover, the variability of orientation representation in visual cortex was tied to the precision of behaviorally probed memory representations (Ester, Anderson, Serences, & Awh, 2013). Memory for more complex visual objects has been evident further upstream in the visual hierarchy (Courtney, Ungerleider, & Haxby, 1996).

Nevertheless, neurophysiology has demonstrated little evidence for persistent neuronal spiking in primary sensory regions such as V1 and motion area MT during memory retention (Bisley, Zaksas, Droll, & Pasternak, 2004; Ferrera, Rudolph, & Maunsell, 1994; Goldman-Rakic, 1995; Lee, Simpson, Logothetis, & Rainer, 2005; Zaksas & Paternak, 2006). Only one study found sustained firing in V1 during a memory delay, but not in the absence of sensory input, as monkeys viewed a background texture throughout (Super, Spekrijse, & Lamme, 2001). Firstly, there is something to be said for not maintaining memories via sustained firing in sensory-dedicated areas: Such spiking could be detrimental for keeping external sensory information separated from internally generated information. Secondly, from the perspective of population coding models the discrepancy between imaging and physiological findings comes as no surprise: Neuroimaging work has decoded mnemonic representations by picking up on population responses, which is perfectly compatible with the absence of sustained spiking in single neurons, and the storage of information in dynamic population codes.

Alternatively, recent work has shown that while sustained spiking was absent from area MT in monkeys remembering motion direction, Local Field Potentials – or LFP's, believed to represent synaptic activity (Mitzdorf, 1985) and closely linked to the BOLD signal measured via fMRI (Logothetis, Pauls, Augath, Trinath, & Oeltermann, 2001) – *did* reflect motion specific information (Mendoza-Halliday, Torres, & M artinez-Trujillo, 2014). This implies that the information present in primary sensory regions during retention might be in the form of LFP's, which could explain why it routinely shows up in fMRI studies measuring BOLD but not in single neuron physiology. This information could be epiphenomenal. For example, due to its coarser spatial extent LFP's might reflect information represented within a neuronal network undergoing dynamic state changes. However, Mendoza-Halliday et al. (2014) linked the information contents of LFP's to increased spiking synchrony between MT and lateral prefrontal cortex, and synchrony in turn was associated with better behavioral performance. Consequently, the author's proposed that early sensory activity is *functionally relevant*, and modulates responses to new sensory inputs (Mendoza-Halliday, et al., 2014; van de Ven, Jacobs, & Sack, 2012).

Indeed, working memory contents can increase cortical excitability (Cattaneo, Pisoni, Papagno, & Silvanto, 2011), bias neuronal firing (Lui & Pasternak, 2011; Miller & Desimone, 1994; Miller, Li, & Desimone, 1991; 1993; Zaksas & Pasternak, 2006;) and fMRI BOLD responses (Sneve, Sreenivasan, Alnæs, Endestad, & Magnussen, 2015). At the behavioral level, the contents of memory can bias perception (Mendoza, Schneiderman, Kaul, & Martinez-Trujillo, 2011; Scocchia et al., 2013) and guide selective attention (Downing, 2000; Soto, Hodsoll, Rotshtein, & Humphreys, 2008). In this sense, working memory acts in a manner that is surprisingly similar to other high-level cognitive contexts. For example, visual mental imagery has been shown to bias perception during binocular rivalry (Pearson, Clifford, & Tong, 2008). Similarly, attention can improve behavioral performance (Carrasco, Ling, & Read, 2004) by enhancing neural processing of attended visual features (Jehee, Brady, & Tong, 2011; Liu et al., 2011; Martinez-Trujillo & Treue, 2004; David, Hayden, Mazer, & Gallant, 2008).

Investigating feature-based attention in visual cortex. As part of this thesis we also attempted to investigate the mechanisms by which feature-based attention mediates behavioral outcomes and perception. Specifically, we looked at orientation processing under conditions of feature-based attention and inattention, by scanning the visual cortex of five healthy participants three times each. First, we wanted to measure orientation tuning at the population level by translating a well-known psychophysical method to the domain of fMRI. This psychophysical method is an orientation-bandpass noise masking procedure, during which participants need to detect a signal (oriented Gabor) within Fourier bandpass filtered noise (Figure 1A). Critically, the orientation contents of the filtered noise can be centered on the orientation of the signal, at which point the signal can be difficult to detect (compare the upper and lower two panels on the far left of Figure 1A), or the orientation contents of the noise can differ from the orientation of the signal. In the latter case, the larger the orientation difference, the easier the signal is to detect (compare the other upper and lower panels in Figure 1A). Psychophysically, the contrast required to detect the signal (contrast threshold) goes down as the signal and noise differ more, and the rate at which it does follows the shape of a half-tuning function, a behavioral proxy for orientation tuning (Figure 1B; and see Experiment 2 in Ling & Blake, 2009).

Our hypothesis was that the psychophysical principles would translate easily to fMRI, and both pilot experiments using a blocked design (for parts of these data see Ling, Pratte, & Tong, 2015), as well as voxel simulations, indicated that BOLD responses and discriminability (d') would be lower when the signal and noise were more similar, and higher as signal and noise would differ more. Thus, our first question was one of methodological innovation, probing a novel way to investigate orientation tuning in early visual cortex. With this hypothetical BOLD derived half-tuning proxy in hand, we would then be able to answer our second question, namely, whether population tuning might sharpen or show gain changes when participants were anticipating the upcoming orientation in a feature-based manner.

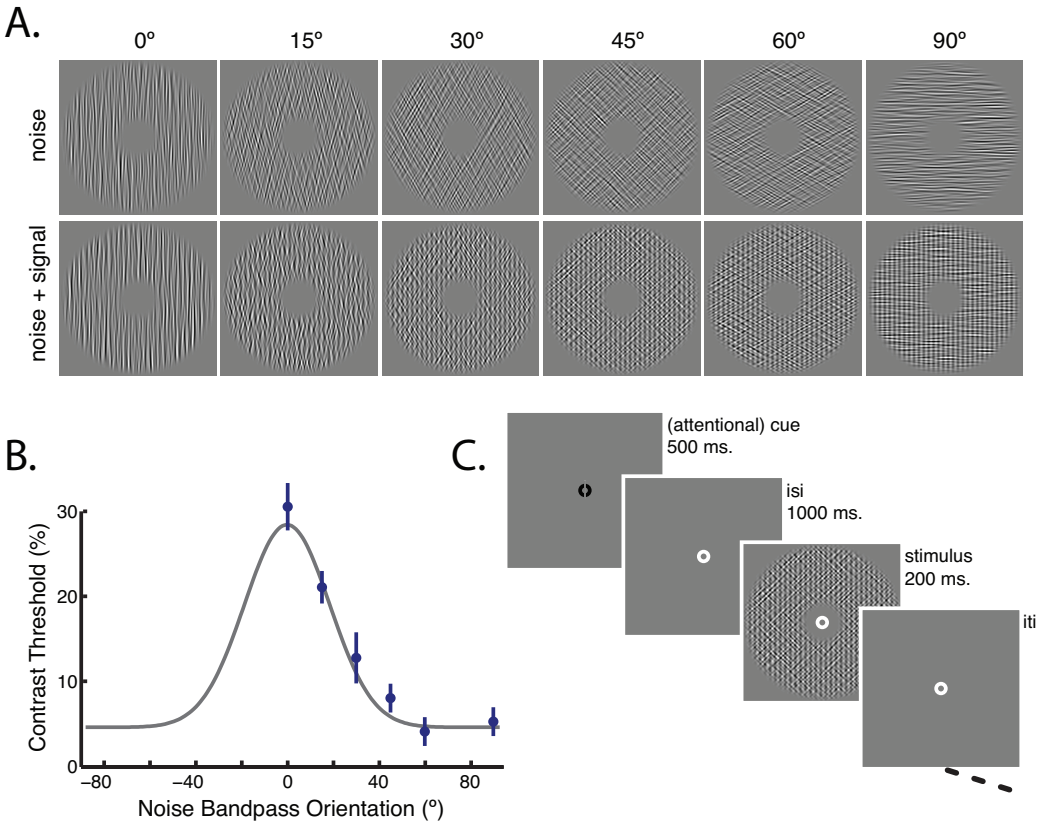


Figure 1. Noise masking procedure **(A)** The top row depicts noise-only stimuli, while the bottom row depicts signal and noise. In these examples the signal is always vertically oriented for illustrative purposes, while in the experiment orientation was randomized. When the signal and noise have the same orientation (far left panels) it is difficult to detect the signal's presence, even knowing it is there. When the signal and noise are orthogonal (far right panels) it's relatively easy to detect the signal. **(B)** Hypothetical contrasts threshold (in blue) plotted as a function of the noise bandpass orientation (assuming a 0º signal). A hypothetical tuning curve is overlaid (in grey) to demonstrate how the noise masking procedure results in a proxy for orientation tuning. **(C)** Participant's task during each of three scanning sessions: A darkening of the fixation (500ms) indicated an upcoming stimulus presentation. On half the trials two small gaps in the fixation indicated the upcoming signal orientation, on the other half of trials no gaps were present. The stimulus appeared for 200ms and participants indicated the presence or absence of a signal via a button-press. The inter-trial interval was 1800, 3800, or 7800ms. Participants performed two localizer runs per session, consisting of the blocked presentation of a flickering checkerboard with the same dimensions as the noise-masking stimuli.

The task we designed (Figure 1C) probed feature-based attention by allowing participants to anticipate the upcoming orientation of the signal on half of the trials. Deployment of

feature-based attention should make it easier for participants to determine whether or not a signal had been embedded in the noise. Stimuli had inner and outer radii of 1.5° and 8° respectively, and the edges were smoothed with a 1° Gaussian kernel (sd = 0.5°). Signal orientation was chosen randomly (1–180°) on each trial, and signal spatial frequency was 2 c/°. The noise had an orientation bandwidth of 5° and a spatial frequency bandwidth of 2 c/°, running between 1 and 4 c/°. The noise was presented at a stable contrast of 40% Michelson, while the signal had a matching or lower contrast. The orientation of the noise relative to that of the signal could be one of six values, as depicted in the columns of Figure 1A.

In order for the feature-based attention manipulation to work, the stimulus presentation had to be very short, or else all behavior would have been at ceiling. This lead us to a fast-event related design, and meant that our BOLD signal was noisy and our beta weights relatively low. When we calculated the differences in beta weights between trials on which the signal was absent and present, along all the signal-noise orientation differences, we were unable to obtain any evidence of tuning. When collapsing across all the signal-noise orientation differences the BOLD response was higher for stimuli with a signal embedded in the noise (p = 0.005) but did not differ with respect to the attentional condition (p = 0.985). Unfortunately, as our novel approach to orientation tuning failed in this fast-event related design, we were also not able to address any questions regarding the deployment of feature-based attention. Other ways of exploration will be needed to further our understanding of how feature-based attention modulates responses to new sensory inputs, and how those behavioral changes might be implemented via tuning changes in human early visual cortex.

Inoculating memories against interference. We have briefly segued from the functional relevance of early sensory involvement in working memory, to the impact of feature-based attention on early visual processing and perception. Here, we will continue with our discussion on working memory, but turn the tables – instead of exploring how memory contents impact perception, we will look at how bottom-up perception might

bias or interfere with representations actively maintained in visual short-term memory. One approach to this question is to look at how memory performance is influenced by interfering information perceived during the delay. Such work has demonstrated that visual memories are affected when the perceived distractors match the memory target on the remembered visual feature (Bennet & Cortese, 1996; Dubé, Zhou, Kahana, & Sekuler, 2014; Huang & Sekuler, 2010; Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Magnussen & Greenlee, 1992; Nemes, Whitaker, Heron, & McKeefry, 2011; Nemes, Parry, Whitaker, & McKeefry, 2012; Rademaker, Bloem, De Weerd, & Sack, 2015; Van der Stigchel, Merten, Meeter, & Theeuwes, 2007). An in-depth discussion of this work on memory interference, and the implications of such work at the representational level, will follow in Chapters 2 and 3 of this thesis.

At the level of implementation, how might high quality memories be achieved in light of constantly interfering information from the eyes? Given the extent of early visual cortex recruitment during memory, and the importance of such areas in bottom-up sensory processing, should a certain sensitivity to interference be expected? Surprisingly little is known about the brain's memory traces in early sensory cortex under the influence of distractors, perhaps due to the typical absence of sustained firing from these areas. Some studies have looked at responses in macaque inferotemporal and prefrontal neurons during the retention of visual stimuli while animals viewed a series of interfering images (Miller et al., 1991; 1993; 1994; 1996). For example, during a match-to-sample task monkeys viewed pictures of common objects, and compared these pictures to a sample maintained in memory (Miller et al., 1991). The majority of cells (85%) exhibited stimulus selective responses, regardless of whether or not a picture matched the memory sample. Other neurons signaled whether or not a picture matched the memory sample, independent of what it depicted (48%). An even smaller proportion was both stimulus and match selective (15%), with stimulus selective firing being attenuated in response to a picture matching the memory sample. The strength of the attenuation response was assumed to signal the degree of similarity between a perceived picture and a memory trace, with higher similarity leading to more attenuation (Miller et al., 1991; 1993). Thus,

different neuronal populations in inferotemporal cortex may serve different mnemonic functions, with some acting as filters that pass only new and unexpected information, presumably due to changes in synaptic weights (Miller et al., 1991; 1993). Later work by the same group showed a second, more active, mnemonic mechanism in the form of response enhancement when the animal was actively searching for an item in memory and detected a match (Miller et al., 1994). While this work elucidates possible components of the memory system, it does little to inform how top-down memory representations interact with bottom-up sensory information, other than assuming different neuronal populations responsible for each. From the behavioral work alluded to above, and discussed in Chapters 2 & 3, we know that bottom-up sensory information *does* have a serious impact on memory representations, and it would be interesting to know at what level of the brain these interactions play out.

The actions of a network. We started our overview of how a visual working memory system might be implemented in the brain by discussing sensory recruitment, and delving into the role of early visual cortex. Now, we will zoom out again and briefly investigate the role of more anterior areas, and prefrontal cortex in particular. What might prefrontal cortex be doing during short-term memory maintenance? Prefrontal activity patterns are well suited for representing information of a high-dimensional nature, and prefrontal cortex is known to encode abstract representations such as arbitrary object categories (Freedman, Riesenhuber, Poggio, & Miller, 2001; 2003). It is also generally believed that prefrontal cortex represents 'rules', and that by sending rule-dependent signals to other parts of the brain it exerts executive control over distributed cognitive processes (Crowe, et al., 2013). The supposed purpose of prefrontal cortex during working memory maintenance is to modulate sensory activity and enhance the selectivity of representations in sensory cortex via feedback signals (Feredoes, Heinen, Weiskopf, Ruff, & Driver, 2011; Lee & D'Esposito, 2012; Miller & Cohen, 2001; Miller, Vytlacil, Fegen, Pradhan, & D'Esposito, 2011; Sreenivasan, 2014; Zanto, Rubens, Thangavel, & Gazzaley, 2011). Indeed, functional connectivity between frontal and sensory regions is important for adequate processing of mnemonic visual information (Chadick & Gazzaley, 2011; Cohen,

Sreenivasan, & D'Esposito, 2012). In terms of implementation, it's been suggested that persistent firing in frontal cortex does not signify the storage of information per se, but rather an attentional signal directed at internal representations maintained in sensory cortices (Sreenivasan et al., 2014).

Note that while a focus on the network helps to flesh out details of implementation, the basic supposition remains identical to that of sensory recruitment. And while sensory recruitment is central to most thinking about how the working memory system might be implemented in the brain, we should nevertheless be careful not to throw the baby out with the bathwater. The fact that feature specific representations exist in sensory cortex during the memory delay does not rule out that frontal cortex plays a role in this as well. Recently, memory for complex multi-colored blobs was decoded from parietal cortex in addition to visual cortex (Christophel, Hebart, & Haynes, 2012). And even more recently information about visual orientation was also decoded from visual, parietal, and even prefrontal areas of the brain (Ester, Sprague, & Serences, 2015). In a similar vein, remembered visual location was reconstructed from information in visual, parietal, and frontal regions (Sprague, Ester, Serences, 2014). These findings signify the diverse roles played by prefrontal cortex during working memory maintenance, and further supports for the idea that representations are stored in population codes, and via the involvement of entire networks.

In this thesis

In this thesis we address a number of questions regarding the maintenance of visual orientation information over a short delay. We want to know whether such short-term representations are robust against interfering information from the eyes, how they decay over time, and what functional role visual cortex plays in their maintenance. Furthermore, behavioral goals and motor actions fundamentally shape how and what we remember. Presumably they even shape every-and-all actions performed by our brains. In this thesis

we will therefore also ask how the brain might flexibly alter the representation of the acting body to include objects that aid the obtainment of behavioral goals.

In Chapter 2 we explore dynamic interactions between bottom-up sensory information and top-down visual memory maintenance. Mnemonic representations of orientation are systematically biased towards interfering orientations presented during the delay. Mnemonic representations also become noisier due to the conflicting nature of interfering orientations, something that has not been demonstrated before. Biases in orientation memory are diminished or abolished when processing of interfering information is altered, respectively, by making the interfering information task relevant or by eliminating it from conscious awareness. In Chapter 3 we expand on the basic finding showing biased memory representations by asking why this is happening. What kind of mechanism might be behind misrepresented orientation information in memory, and what can the influence of a distractor teach us about the computations required to perform a simple short-term memory task?

While the first two chapters look at the vulnerability of mnemonic representations in the face of interference, Chapter 4 looks at the vulnerability of memory representations as a function of time. The fidelity of memories for several visual features is investigated over variable delays. Similar to the effect of interference, mnemonic representations also prove susceptible to the passage of time. In Chapter 5 we shift the question of visual mnemonic representations to the level of implementation. We ask what the neural mechanisms are by which information is retained over brief intervals, and what the involvement is of visual sensory cortex. By applying TMS over visual cortex during the delay we look into the exact role played by this part of the brain's working memory system.

We switch gears in Chapter 6, and delve into the output side of things. After all, behavioral output is the ultimate means by which living things achieve their ends. Humans in particular have evolved to achieve many more ends than most species through

their extraordinary ability to use tools. While many animals have this ability to some degree, humans are unique in the extent of their skill, allowing them to vastly expand the range of otherwise possible behavioral output. This final empirical chapter demonstrates the flexible nature of this skill, and how the body's representation can be adapted to include tools in order to attain ever more sophisticated behavioral goals.

References

Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., & de Lange, F.P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431.

Alvarez, G.A., & Cavanagh, P. (2004). The capacity of visual short-term memory is set both by visual information load and by number of objects. *Psychological Science*, *15*(2), 106–111.

Baddeley, A.D. & Hitch, G.J. (1974). Recent Advances in Learning and Motivation, Academic: New York, 47–89.

Barak, O., Tsodyks, M., & Romo, R. (2010). Neuronal population coding of parametric working memory. The Journal of Neuroscience, 30(28), 9424–9430.

Bays, P.M. (2014). Noise in Neural Populations Accounts for Errors in Working Memory. *Journal of Neuroscience*, *34*(10), 3632–3645.

Bays, P.M. (2015). Spikes not slots: noise in neural populations limits working memory. *Trends in Cognitive Sciences*, *19*(8), 431–438.

Bays, P.M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Bays, P.M., Catalao, R.F.G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11.

Bennett, P.J. & Cortese, F (1996). Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. *Vision Research, 36*(2), 233–238.

Bisley, J.W., Zaksas, D., Droll, J.A., & Pasternak, T. (2004). Activity of neurons in cortical area MT during a memory for motion task. *Journal of Neurophysiology*, *91*(1), 286–300.

Blake, R., Cepeda, N.J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 353.

Bosch, S.E., Jehee, J.F.M., Fernández, G., & Doeller, C.F. (2014). Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *The Journal of Neuroscience*, *34*(22), 7493–7500.

Braitenberg, V., & Schüz, A. (1991). Anatomy of the Cortex – Statistics and Geometry. Vol 18. Springer Verlag: Heidelberg, Germany.

Burgess, G.C, Gray, J.R., Conway, A.R.A., & Braver, T.S. (2011) Neural mechanisms of interference control underlie the relationship between fluid intelligence and working memory span. *Journal of Experimental Psychology General, 140*(4), 674–692.

Burnett Heyes, S., Zokaei, N., Van Der Staaij, I., Bays, P.M., & Husain, M. (2012). Development of visual working memory precision in childhood. *Developmental Science*, *15*(4), 528–539.

Buschman, T.J., Siegel, M., Roy, J.E., & Miller, E.K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11252–11255.

Carandini, M., & Heeger, D.J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62.

Carrasco, M., Ling, S., & Read, S. (2004). Attention alters appearance. *Nature Neuroscience*, *7*(3), 308–313.

Cattaneo, Z., Vecchi, T., Pascual-Leone, A., & Silvanto, J. (2009). Contrasting early visual cortical activation states causally involved in visual imagery and short-term memory. *European Journal of Neuroscience, 30*(7), 1393–1400.

Cattaneo, Z., Pisoni, A., Papagno, C., & Silvanto, J. (2011). Modulation of Visual Cortical Excitability by Working Memory: Effect of Luminance Contrast of Mental Imagery. *Frontiers in Psychology*, *2*, 1–9.

Chadick, J.Z. and Gazzaley, A. (2011) Differential coupling of visual cortex with default or frontal-parietal network based on goals. *Nature Neuroscience, 14*, 830–832.

Cohen, J.R., Sreenivasan, K.K., & D'Esposito, M. (2012). Correspondence between stimulus encoding- and maintenance-related neural processes underlies successful working memory. *Cerebral Cortex, 24*, 593–599.

Courtney, S.M., Ungerleider, L.G., Keil, K., & Haxby, J.V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex, 6*(1), 39–49.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–185.

Crowder, R.G. (1993). Auditory memory. In S. McAdams & E. Bigand (Eds.), Thinking in sound (pp. 113-145). Oxford, England: Oxford press.

Crowe, D.A., Goodwin, S.J., Blackman, R.K., Sakellaridi, S., Sponheim, S.R., MacDonald, A.W., & Chafee, M.V. (2013). Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nature Neuroscience, 16*(10), 1484–1491.

David, S.V., Hayden, B.Y., Mazer, J.A., & Gallant, J.L. (2008). Attention to stimulus features shifts spectral tuning of V4 neurons during natural vision. *Neuron*, *59*(3), 509–521.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473.

Dubé, C., Zhou, F., Kahana, M.J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research*, *96*, 8–16.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegnér, J., & Compte, A. (2009). Mechanism for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences of the United States of*

America, *106*(16), 6802–6807.

Ester, E.F., Anderson, D.E., Serences, J.T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience*, *25*(5), 754–761.

Ester, E.F., Serences, J.T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *The Journal of Neuroscience, 29*(48), 15258–15265.

Feredoes, E., Heinen, K., Weiskopf, N., Ruff, C., & Driver, J. (2011). Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proceedings of the National Academy of Sciences of the United States of America, 108*(42), 17510–17515.

Ferrera, V.P., Rudolph, K.K., & Maunsell, J.H. (1994). Responses of neurons in the parietal and temporal visual pathways during a motion task. *The Journal of Neuroscience*, *14*(10), 6171–6186.

Fougnie, D., Suchow, J.W., & Alvarez, G.A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229.

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science, 291*, 312–316

Freedman, D.J., Riesenhuber, M., Poggio, T., Miller, E.K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience, 23*(12), 5235–5246.

Fukuda, K., Vogel, E., Mayr, U., & Awh, E. (2010). Quantity, not quality: the relationship between fluid intelligence and working memory capacity. *Psychonomic Bulletin & Review*, *17*(5), 673–679.

Funahashi, S., Bruce, C. J., and Goldman-Rakic, P. S. (1989). Mnemonic coding of visual space in the primate dorsolateral prefrontal cortex. *Journal of Neurophysiology, 61*, 331-349.

Fuster, J.M., & Jervey, J.P. (1981). Inferotemporal neurons distinguish and retain behaviorally relevant features of visual stimuli. *Science, 212*(4497), 952–955.

Fuster, J.M., & Jervey, J.P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. *The Journal of Neuroscience, 2(3), 361-375.*

Goldman-Rakic, P.S. (1995) Cellular basis of working memory. *Neuron, 14*, 477–485.

Georgopoulos, A.P., Schwartz, A.B., & Kettner, R.E. (1986). Neuronal population coding of movement direction. *Science, 233*, 1416–1419.

Graf, A.B.A., Kohn, A., Jazayeri, M., & Movshon, J.A. (2011). Decoding the activity of neuronal populations in macaque primary visual cortex. *Nature Neuroscience*, *14*(2), 239–245.

Gruesser, O-J. & Gruesser-Cornehls, U. (1978). Physiology of vision. In Fundamentals of sensoruy physiology, New York, Springer-Verlag.

Harris, J.A., Miniussi, C., Harris, I.M., & Diamond, M.E. (2002). Transient storage of a tactile memory trace in primary somatosensory cortex. *The Journal of Neuroscience*, *22*(19), 8720–8725.

Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–197.

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, *340*(6132), 639–642.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: two classes of attractors at work. *Journal of Vision*, *10*(2):24, 1–27.

James, W. (1890). The principles of psychology. New York: H. Holt and Company.

Jehee, J.F.M., Brady, D.K., & Tong, F. (2011). Attention improves encoding of task-relevant features in the human visual cortex. *The Journal of Neuroscience, 31*(22), 8210–8219.

Johnson, M.K., McMahon, R.P., Robinson, B.M., Harvey, A.N., Hahn, B., Leonard, C.J., Luck, S.J., & Gold, J.M. (2013). The relationship between working memory capacity and broad measures of cognitive ability in healthy adults and people with schizophrenia. *Neuropsychology, 27*(2), 220–229.

Kane, M.J., and Engle, R.W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: an individual-differences perspective. *Psychonomic Bulletin & Review, 9*, 637–671.

Kaufman, E.L., Lord, M.W., Reese, T.W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology, 62*(4), 498–525.

Kennedy, H., Knoblauch, K., & Toroczkai, Z. (2013). Data coherence and completion actually do count for interareal cortical network. *Neuroimage 80*, 37–45.

Kinchla, R.A., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception & Psychophysics 2*(6), 219–229.

Kosslyn, S.M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J.P., Thompson, W.L., Ganis, K., Sukel, E., & Alpert, N.M. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science*, *284*(5411), 167–170.

Lee, B., & Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Research*, *36*(14), 2159–2166.

Lee, S.-H., Kravitz, D.J., & Baker, C.I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073.

Lee, T.G. & D'Esposito, M. (2012) The dynamic nature of top-down signals originating from prefrontal cortex: a combined fMRI–TMS study. *The Journal of Neuroscience, 32*, 15458–15466.

Lee, H., Simpson, G.V., Logothetis, N.K., & Rainer, G. (2005). Phase locking of single neuron activity to theta oscillations during working memory in monkey extrastriate visual cortex. *Neuron*, *45*(1), 147–156.

Ling, S., & Blake, R. (2009). Suppression during binocular rivalry broadens orientation tuning. *Psychological*, *20*(11), 1348–1355.

Ling, S., Pratte, M.S., & Tong, F. (2015). Attention alters orientation processing in the human lateral geniculate nucleus. *Nature Neuroscience,* advance online publication.

Liu, T., Hospadaruk, L., Zhu, D.C., & Gardner, J.L. (2011). Feature-specific attentional priority signals in human cortex. *The Journal of Neuroscience*, *31*(12), 4484–4495.

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Lui, L.L., & Pasternak, T. (2011). Representation of comparison signals in cortical area MT during a delayed direction discrimination task. *Journal of Neurophysiology*, *106*(3), 1260–1273.

Magnussen, S., Greenlee, M.W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, *31*(7-8), 1213–1219.

Magnussen, S., & Greenlee, M.W. (1992). Retention and disruption of motion information in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 151–156.

Marr, D. (1982). Vision: A computational investigation into the human representation and processing of visual information. W.H. Freeman, New York.

Martinez-Trujillo, J.C., & Treue, S. (2004). Feature-based attention increases the selectivity of population responses in primate visual cortex. *Current Biology*, *14*(9), 744–751.

Mendoza, D., Schneiderman, M., Kaul, C., & Martinez-Trujillo, J. (2011). Combined effects of feature-based working memory and feature-based attention on the perception of visual motion direction. *Journal of Vision*, *11*(1):11, 1–15.

Mendoza-Halliday, D., Torres, S., & Martinez-Trujillo, J.C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature Neuroscience*, *17*(9), 1255–1262.

Meyers, E.M., Freedman, D.J., Kreiman, G., Miller, E.K., & Poggio, T. (2008). Dynamic population coding of category information in inferior temporal and prefrontal cortex. *Journal of Neurophysiology, 100*, 1407–1419.

Meyers, E.M. Qi, X.-L., & Constantinidis, C. (2012). Incorporation of new information into prefrontal cortical activity after learning working memory tasks. *Proceedings of the National Academy of Sciences of the United States of America, 109*(12), 4651–4656.

Meyer, K., Kaplan, J.T., Essex, R., Webber, C., Damasio, H., & Damasio, A. (2010). Predicting visual stimuli on the basis of activity in auditory cortices. *Nature Neuroscience*, *13*(6), 667–668.

Miller, B.T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito (2011). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience, 23*(1), 1–10.

Miller, E.K. and Cohen, J.D. (2001) An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24*, 167–202.

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, *263*(5146), 520–522.

Miller, E.K., Erickson, C.A., & Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *The Journal of Neuroscience*, *16*(16), 5154–5167.

Miller, E.K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*, *254*(5036), 1377–1379.

Miller, E.K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience*, *13*(4), 1460–1478.

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic theory of working memory. *Science, 319*, 1543–1546.

Nemes, V. A., Parry, N. R., Whitaker, D., & McKeefry, D. J. (2012). The retention and disruption of color information in human short-term visual memory. *Journal of Vision*, *12*(1):26, 1–14.

Nemes, V. A., Whitaker, D., Heron, J., & McKeefry, D. J. (2011). Multiple spatial frequency channels in human visual perceptual memory. *Vision Research*, *51*(23), 2331–2339.

Nyberg, L., Habib, R., McIntosh, A.R., & Tulving, E. (2000). Reactivation of encoding-related brain activity during memory retrieval. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11120–11124.

O'Craven, K., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

Pearson, B., Raškevičius, J., Bays, P.M., Pertzov, Y., & Husain, M. (2014). Working memory retrieval as a decision process. *Journal of Vision*, *14*(2):2, 1–15.

Pearson, J., Clifford, C.W.G., & Tong, F. (2008). The Functional Impact of Mental Imagery on Conscious Perception. *Current Biology*, *18*(13), 982–986.

Pearson, J., Rademaker, R.L., & Tong, F. (2011). Evaluating the Mind's Eye: The Metacognition of Visual Imagery. *Psychological Science*, *22*(12), 1535–1542.

Phillips, W.A. (1974) On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*(2), 283–290.

Pylyshyn, Z.W., & Storm, R.W. (1988). Tracking multiple independent targets: evidence for a parallel tracking mechanism. *Spatial Vision*, *3*(3), 179–197.

Rademaker, R.L., Tredway, C.H., & Tong, F. (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision*, *12*(13), 21–21.

Rademaker, R.L., & Pearson, J. (2012). Training visual imagery: improvements of metacognition, but not imagery strength. *Frontiers in Psychology*, *3*(224), 1–11.

Rademaker, R.L., Bloem, I.M., De Weerd, P., & Sack, A.T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance*. Advance online publication.

Ranganath, C. and D'Esposito, M. (2005) Directing the mind's eye: prefrontal, inferior and medial temporal mechanisms for visual working memory. *Current Opinion in Neurobiology, 15*, 175–182.

Rayner, K. (1978). Eye movements in reading and information processing. *Psychological bulletin, 85(3)*, 618–660.

Regan, D. (1985). Storage of spatial-frequency information and spatial-frequency discrimination. *Journal of the Optical Society of America, 2*(4), 619–621.

Riggall, A.C., & Postle, B.R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *The Journal of Neuroscience*, *32*(38), 12990–12998.

Schüz, A., & Braitenberg, V. (2002). The human cortical white matter: quantitative aspects of cortico-cortical long-range connectivity. In: Cortical areas: unity and diversity. London, Taylor & Francis.

Scocchia, L., Cicchini, G.M., & Triesch, J. (2013). What's 'up'? Working memory contents can bias orientation processing. *Vision Research*, *78*, 46–55.

Serences, J.T., Ester, E.F., Vogel, E.K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214.

Shafi, M., Zhou, Y., Quintana, J., Chow, C., Fuster, J., Bodner, M. (2007). Variability in neuronal activity in primate cortex during working memory tasks. *Neuroscience, 146(3)*, 1082–1108.

Silvanto, J. & Soto, D. (2011). Causal evidence for subliminal percept-to-memory interference in early visual cortex. *Neuroimage, 59*, 840–845.

Simons, D.J., & Levin, D.T. (1997). Change Blindness. *Trends in Cognitive Sciences, 1*(7), 261–267.

Sligte, I.G., Scholte, S.H., & Lamme, V.A.F. (2008). Are there multiple visual short-term memory stores? *PLoS ONE 3*:e1699.

Sligte, I.G., Scholte, S.H., & Lamme, V.A.F. (2009). V4 activity predicts the strengths of visual short-term memory representations. *The Journal of Neuroscience, 29*(23), 7432–7438.

Sneve, M.H., Sreenivasan, K.K., Alnæs, D., Endestad, T., & Magnussen, S. (2015). Short-term retention of visual information: Evidence in support of feature-based attention as an underlying mechanism. *Neuropsychologia*, *66*(C), 1–9.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G.W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*(9), 342–348.

Sperling, G., (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1–29.

Sprague, T.C., Ester, E.F., & Serences, J.T. (2014). Reconstructions of Informationin Visual Spatial Working Memory Degrade with Memory Load. *Current Biology*, *24*(18), 2174–2180.

Sreenivasan, K.K., Curtis, C.E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89.

Stokes M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron, 78*(2), 364–375.

Stokes, M.G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic coding for cognitive control in prefrontal cortex. *Neuron*, *78*(2), 364–375.

Supèr, H., Spekreijse, H., & Lamme, V.A. (2001). A neural correlate of working memory in the monkey primary visual cortex. *Science*, *293*(5527), 120–124.

Trick, L.M., & Pylyshyn, Z.W. (1994). Why are small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review*, *101*(1), 80–102.

van de Ven, V., Jacobs, C., & Sack, A.T. (2012). Topographic Contribution of Early Visual Cortex to Short-Term Memory Consolidation: A Transcranial Magnetic Stimulation Study. *The Journal of Neuroscience*, *32*(1), 4–11.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W.J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785.

Van der Stigchel, S., Merten, H., Meeter, M., & Theeuwes, J. (2007). The effects of a task-irrelevant visual event on spatial working memory. *Psychonomic Bulletin & Review*, *14*(6), 1066–1071.

Vogel, E.K., Woodman, G.F., & Luck, S.J. (2001). Storage of features, con- junctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance, 27*, 92–114.

Watanabe K., & Funahashi S. (2014). Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. *Nature Neuroscience, 17*, 601–611.

Wheeler, M.E., Petersen, S.E., & Buckner, R.L. (2000). Memory's Echo: Vivid Remembering Reactivates Sensory-Specific Cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(20), 11125–11129.

Wilken, P., & Ma, W. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12).

Zaksas, D., & Pasternak, T. (2006). Directional Signals in the Prefrontal Cortex and in Area MT during a Working Memory for Visual Motion Task. *The Journal of Neuroscience*, *26*(45), 11726–11742.

Zanto, T.P., Rubens, M.T., Thangavel, A., & Gazzeley, A. (2011) Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience, 14*(5), 656–661.

Zhang, W., & Luck, S.J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.

Chapter 2

The impact of interference on short-term memory for visual orientation

Corresponding publication:

Rademaker R.L., Bloem I.M., De Weerd P., Sack A.T. (2015).

The impact of interference on short-term memory for visual orientation.

Journal of Experimental Psychology: Human Perception and Performance. Advance online publication.

Abstract

Visual short-term memory serves as an efficient buffer for maintaining no longer directly accessible information. How robust are visual memories against interference? Memory for simple visual features has proven vulnerable to distractors containing conflicting information along the relevant stimulus dimension, leading to the idea that interacting feature-specific channels at an early stage of visual processing support memory for simple visual features. Here we showed that memory for a single randomly orientated grating was susceptible to interference from a to-be-ignored distractor grating presented midway through a 3s delay period. Memory for the initially presented orientation became noisier when it differed from the distractor orientation, and response distributions were shifted *towards* the distractor orientation (by ~3°). Interestingly, when the distractor was rendered task-relevant by making it a second memory target, memory for both retained orientations showed reduced reliability as a function of increased orientation differences between them. However, the degree to which responses to the first grating shifted towards the orientation of the task-relevant second grating was much reduced. Finally, using a dichoptic display, we demonstrated that these systematic biases caused by a consciously perceived distractor disappeared once the distractor was presented outside of participants' awareness. Together, our results show that visual short-term memory for orientation can be systematically biased by interfering information that is consciously perceived.

Introduction

Visual perception is a highly complex process aimed at making sense of a dynamic external world by constructing a coherent visual percept from rapidly changing retinal images. To keep visual information online in order to perform necessary computations, the brain needs to maintain this information after it can no longer be sensed directly, and inoculate it against interference from other inputs. Visual 'short-term', or 'working' memory serves as a highly efficient buffer that temporarily stores sensory information for future use. Research aiming to better characterize visual working memory often focuses on one particularly captivating feature of the system, which is its profound capacity limitation, comprised of about 3-4 items (Bays & Husain, 2008; Fougnie, Suchow, & Alvarez, 2012; Fukuda, Awh, & Vogel, 2010; Luck & Vogel, 1997; Palmer, 1990; van den Berg, Shin, Chou, George, & Ma, 2012; Wilken & Ma, 2004; Zhang & Luck, 2008). However, an emphasis on *quantity* is often confounded with limitations equally applicable to encoding and perception in general (Gazzaley & Nobre, 2012; Palmer, 1990), and does little to inform memory *quality*. What happens to the quality of visual memories once they have been well and truly transferred into internal representations?

Traditionally, research into memory quality falls under the label 'short-term memory'. A core question concerns the extent to which new information from the eyes has the potency to interfere with information already in memory. One way to investigate this is by keeping the stimuli at the encoding stage constant, as well as keeping set size within the confines of supposed (cognitive) capacity limitations. Doing precisely this, early psychophysical work into visual short-term memory has shown that memories are not immune against interference from other stimuli: When people remembered a single spatial frequency (Bennett & Cortese, 1996; Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Nemes, Whitaker, Heron, & McKeefry, 2011), velocity (Magnussen & Greenlee, 1992; McKeefry, Burton, & Vakrou, 2007; Pasternak & Zaksas, 2003), or color (Nemes, Parry, Whitaker, & McKeefry, 2012; Nilsson & Nelson, 1981), a second irrelevant stimulus shown during the retention interval degraded performance on a delayed discrimination task. Interference

effects (also referred to as 'memory masking') have also been found when a to-be-ignored stimulus was presented shortly before the memory task (Lalonde & Chaudhuri, 2002), and in other modalities such as short-term memory for pitch (Deutsch, 1970; 1973).

Visual memory for simple features proved vulnerable only to distractors containing conflicting information along the relevant stimulus dimension (Magnussen et al., 1991; Magnussen & Greenlee, 1992; Magnussen & Greenlee 1997). For example, in the case of a remembered spatial frequency this meant that an irrelevant but different spatial frequency affected thresholds, whereas an irrelevant orientation change –without a change in spatial frequency– did not (Lalonde & Chaudhuri, 2002; Magnussen et al., 1991; Nemes et al., 2011). The fact that interference acts independently on low-level features implicates higher-level visual areas beyond primary visual area V1 as a locus of interference (Magnussen, 2000; Magnussen & Greenlee, 1999). This idea is further supported by findings demonstrating that interference obeyed size constancy (Bennett & Cortese, 1996), and was location invariant (Nemes et al., 2011; 2012; Phillips, 1974), still occurring under free-viewing conditions (Magnussen et al., 1991). Only at higher levels of representation are visual features, size, and retinal location processed independently. Recent neuroimaging work confirms the likely locus of interference at a higher level of the visual hierarchy, strongly implicating V4 in particular (Sneve, Sreenivasan, Alnæs, Endestad, & Magnussen, 2015).

These findings have spawned the idea that memories are stored in narrowly tuned feature-specific channels in visual cortex, where inhibitory cross-channel interactions are responsible for the psychophysically observed distractor effects (Magnussen, 2000; Magnussen & Greenlee, 1992; 1999; Nemes et al., 2011). In this view, information loss is due to inhibition between different memory stores maintaining conflicting information of a shared visual feature, such as two spatial frequencies that differ by one octave (Magnussen et al., 1991). This idea aligns with the observation that an irrelevant stimulus did not impact performance when it matched the memorized stimulus on the task relevant feature (Lalonde & Chaudhuri, 2002; Magnussen et al., 1991; Magnussen &

Greenlee, 1992), while the most prevalent disruption occurred when a distractor differed from the memory target by one octave of more (in the case of spatial frequency), or by about twice the velocity (in the case of velocity).

More recently it was discovered that the deleterious effects of a distractor on memory did not result from a drop in memory fidelity, but from an attraction of representations in memory towards the distractor (Huang & Sekuler, 2010a). When participants viewed two subsequently presented gratings of different spatial frequencies, having to report only one of them, it was found that the reported spatial frequency was biased towards the spatial frequency of the unreported grating, while the variability of report remained unchanged (Huang & Sekuler, 2010a). Earlier work relying on delayed discrimination tasks had been unable to uncover memory attraction – as such tasks index memory quality or fidelity by a single threshold measure. Instead, this study employed a method of adjustment procedure, allowing a measure of response variability as well as a measure of the response mean (or central tendency). Critically, attraction was stronger when two spatial frequencies were task relevant compared to when one of the two was irrelevant and could be ignored (Dubé, Zhou, Kahana, & Sekuler, 2014; Huang & Sekuler, 2010a).

Based on these findings a modified version of the channel interaction account emerged, in which a second stimulus exerts its influence at a visual stage of processing via (incomplete) perceptual averaging. Such averaging presumably occurs in a population of spatial frequency selective mechanisms. In this view, the degree to which two stimuli are averaged depends on their respective weights, and selective attention might alter these weights such that the influence of irrelevant information can be mostly filtered out (Dubé et al., 2014; Huang & Sekuler, 2010a). A number of key predictions follow from this account: The more a distractor differs from an item in memory, the more strongly it will attract memory due to averaging. Attraction should be reduced for irrelevant and unattended distractors by means of attentional filtering. It's furthermore implied that the influence of the second grating will always be one of attraction, as illustrated by a magnet metaphor (Huang & Sekuler, 2010a).

While attraction has been shown in the context of spatial frequency (Dubé et al., 2014; Huang & Sekuler, 2010a; Nemes et al., 2011) and color (Nemes et al., 2012), it remains unclear whether and how a distractor might bias memory representations of orientation. Both short- and long-term memory for orientation, draw upon visual cortical regions (Bosch, Jehee, Fernández, & Doeller, 2014; Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009; Sneve, Alnæs, Endestad, Greenlee, & Magnussen, 2012) making it likely that orientation, like other low-level features, is susceptible to interference. Such susceptibility should come as no surprise, considering that interfering visual information necessarily enters the same sensory regions as those responsible for memory maintenance.

We designed our study to investigate memory interference for orientation under a variety of circumstances, testing the channel interaction (Magnussen, 2000) and perceptual averaging (Dubé et al., 2014; Huang & Sekuler, 2010a) theories proposed by previous work. First, it is currently not known whether memory biases other than attraction exist. Specifically, changes in memory variability or 'noise' due to a distractor have not been demonstrated before, but an increase in noise logically follows if one adopts the channel interaction account (Magnussen, 2000). Conversely theories such as optimal cue integration (Ernst & Banks, 2002) predict a noise reduction, since integration of the distractor would presumably cause a decrease in variance. Another open question is whether or not attraction might depend on the range of memory target and distractor differences used within a single experiment. We explored whether attraction might become stronger, returns to baseline, or scales with a larger range of target-distractor differences (Experiments 1 and 2), as the current evidence is conflicting (Huang & Sekuler, 2010a; Magnussen et al., 1991; Magnussen & Greenlee, 1992; Nemes et al., 2011; 2012). We furthermore contrasted interference from both task relevant and irrelevant distractors (Experiments 1 and 2 versus Experiment 3). A perceptual averaging mechanism (Dubé et al., 2014; Huang & Sekuler, 2010a) assumes less interference from task irrelevant information, as attention acts to largely filter out a distractor's influence. Conversely, the channel interaction account (Magnussen, 2000) is agnostic to the relevance of the interfering information. How well these accounts hold for orientation memory remains to

be seen. Finally, to address why irrelevant information would be integrated in the first place we looked at whether mere bottom-up information that is not processed consciously would be sufficient to interfere with information in memory (Experiment 4).

To answer these questions it was vital to parse more general performance changes into independent contributions. By using the method of adjustment we were able to construct error distributions, the shape of which can disclose fundamental mechanisms behind changes in memory performance (Ma, Husain, & Bays, 2014). For example, studies quantifying working memory limitations routinely rely on error distributions to infer information about whether or not an item is in memory (Wilken & Ma, 2004; Zhang & Luck, 2008), whether people might inadvertently report the wrong item (Bays, Catalao, & Husain, 2009), or whether memories are variable from trial-to-trial and item-to-item (Fougnie et al., 2012; van den Berg et al., 2012). Here we exploit this methodology to investigate memory quality, fitting a circular Gaussian (or 'von Mises') distribution to retrieve the noisiness of memory (indexed by the standard deviation, or *SD*) and the distribution mean (μ), allowing a more comprehensive insight into the dynamics underlying memory interference. Additionally, because we expected memory attraction to shift the distribution mean, we compared the fits from a von Mises with and without a parameter describing the distribution mean. Such a comparison provides us with additional information, namely, instead of examining whether a shift exists, it assesses whether assuming such a shift helps explain the data.

Here we followed the psychophysical short-term memory tradition of presenting only a single stimulus that is then translated into a high fidelity memory, and looked at its robustness against interference. We showed that short-term memory for orientation is not immune against interference, and systematic biases emerge when a distractor orientation is presented that differs from an orientation held in memory. These biases required awareness and consisted of increased memory noise, which has not been demonstrated before, and a shift in the response distribution towards the distractor orientation or 'attraction'. The range of target-distractor differences did not impact these biases, while

attention reduced memory attraction, even leading to occasional instances of memory repulsion. These findings are not predicted by previous theories, and require a revision of current models on memory quality and interference.

Experiment 1

Methods

Participants. Eight healthy volunteers (6 female) between the ages of 21 and 29 (M = 25.38; SE = 1.12) participated in Experiment 1. All participants had normal or corrected-to-normal vision, and provided informed consent. The study took place under the approval of the standing ethical committee of the Psychology and Neuroscience department at Maastricht University. With the exception of two of the authors (RR and IB), participants received monetary reimbursement for their time and were naïve to the purpose of the study.

Stimuli. All experimental stimuli were viewed in a dark room on a luminance-calibrated CRT monitor with 1280 x 1024 resolution and 60 Hz refresh rate. Visual stimuli were generated using MATLAB 7.5.0 (R2007b) and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) on a PC running windows XP. Stimuli consisted of centrally presented oriented gratings with a spatial frequency of 2 c/°, and a diameter of 3° of visual angle. Gratings were presented at 20% Michelson contrast with added jitter (randomly selected from a uniform distribution with a range spanning $\pm$ 10% contrast), within a wide Gaussian envelope (sd = 2.5°) on a uniform grey background that shared the same mean luminance of 40.8 cd/m^2. Grating phase was randomized between 0 and 2π. The test stimulus used to obtain participants' responses was a mouse-probe consisting of the centrally presented white bull's eye fixation (0.5° of visual angle in diameter) and an interrupted white line, of which each segment was 0.025° wide and 0.125° long. The two line segments were spaced 3° apart to ensure that their visual field position was non-

overlapping with that of the previously presented grating. By moving the mouse around, the interrupted white line rotated about the fixation bull's eye, allowing participants to replicate the orientation in memory by method-of-adjustment. Participants were seated at a viewing distance of 57 cm, and a chinrest assisted in maintaining head stability. Participants were instructed to maintain steady fixation throughout all experimental trials.

Procedure. Throughout all the experiments described here, the general outline of the task was the same (see Figure 1 for reference). First, a target grating with a randomly chosen orientation between 1 and 180° was presented for 200 ms, and participants remembered the orientation of this grating. After the retention-interval a test (mouse-probe) was presented at an initially random orientation also between 1 and 180°, and participants rotated this dial to match the orientation in memory. Once a participant was satisfied with the response, a left mouse-click allowed them to continue to the next trial. Precision of replication-performance was stressed throughout all experiments described in this paper, and there were never any time constraints for participants' responses.

We first established a baseline performance for this particular method-of-adjustment probe when a single orientation was memorized. In two separate blocks of 200 trials each, participants remembered a randomly oriented grating for 1, 3, 6, or 12 seconds (randomly interleaved), after which they rotated the dial to match the orientation in memory as closely as possible. The trials with a retention interval of 3-seconds were subsequently used as the 'no distractor' baseline for the rest of the experiment.

For the main part of the experiment a second distractor grating was introduced, and presented for 200 ms halfway through a fixed 3–second retention interval (Figure 1). The orientation of the distractor could be one of several orientations that were fixed relative to the target. These relative orientations were sampled in a Gaussian (δ=25) fashion, resulting in denser sampling around the target orientation. The distractor orientation could differ –15°, –7°, –4°, –2°, 0°, 2°, 4°, 7° or 15° from the randomly chosen target orientation. Thus, the distractor grating could appear at an orientation that was

counterclockwise, the same, or clockwise relative to the target (randomly interleaved). Participants were told that the second grating was completely irrelevant to the task, and were instructed to ignore it. This part of the experiment consisted of 900 trials in total, divided over 5 blocks (~22 minutes per block). Participants were allowed a short practice before the start of the experiment, and while half of them started with the two baseline blocks, the other half started with the main experiment.

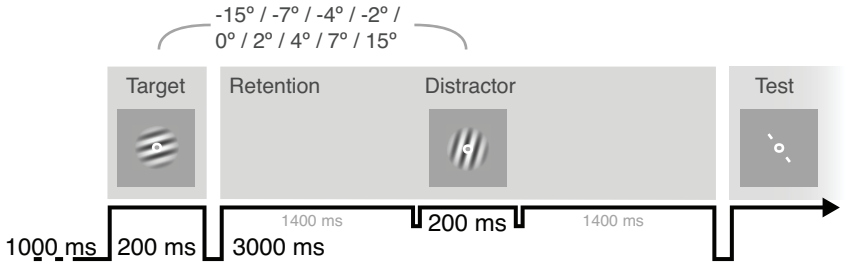


Figure 1. Trial sequence Experiment 1. Participants viewed a randomly orientated grating for 200 ms at the start of each trial. They retained the orientation in memory over a 3-second interval while fixating a white bull's eye presented against a mean-grey background. Midway through the retention interval, a second, to-be-ignored grating was presented for 200ms. The orientation of this to-be-ignored distractor grating could be the same as the orientation in memory, or it could be rotated 2, 4, 7, or 15 degrees counter-clockwise or clockwise relative to the orientation in memory (randomly interleaved). After the retention interval participants were presented with a test stimulus, which they could rotate by using the computer mouse to match the orientation in memory as precisely as possible. When satisfied with their response, participants clicked the mouse and continued to the next trial 1 second later.

Analyses. For each condition of interest a distribution of response errors was obtained by calculating the difference between target and response (reported orientation minus target orientation). Memory accuracy is the average (absolute) orientation-error in each condition. In order to look beyond simple accuracies, and to take the entire response distribution into account, we also estimated relevant characteristics from these response distributions by fitting a von Mises function (circular analog of a normal distribution) to the response distributions for the experiments described in this paper. A von Mises describes the data in terms of the mean (μ) and circular variance (*SD*).

Data analysis were performed in MATLAB using custom functions as well as functions provided by the MemToolbox (Suchow, Brady, Fougnie, & Alvarez, 2013), and the Circular Statistics Toolbox (Berens, 2009). Here, we used maximum likelihood estimation to obtain estimates for each parameter value (on which we performed regular repeated-measures statistics). Additionally, we used the Bayesian Information Criterion (BIC, Schwarz, 1978) to compare models with and without a distribution mean as a free parameter.

Results and Discussion

We tested whether an irrelevant distractor presented during retention could systematically affect memory of a single target by parametrically varying the orientation of the distractor relative to the orientation of the target. A within-subject ANOVA revealed that the absolute response error was not the same across the relative orientation differences between the target and distractor gratings (Figure 2A; $F_{(8,56)} = 2.286$; $p = 0.034$). As can also be seen in Figure 2A, this trend was quadratic ($F_{(1,7)} = 10.559$; $p = 0.014$) indicating that larger relative orientation differences between target and distractor led to bigger performance decrements, compared to smaller relative differences. This quadratic effect is reflected by the characteristic 'v-shape' in the data. Additional post-hoc tests (paired *t*-tests) did not reveal a difference between the no-distractor baseline and any of the target-distractor conditions (all $p > 0.149$), implying that the quadratic effect is a mixture of improved memory performance at small, and impaired memory performance at large target-distractor differences.

To examine why performance suffered when the distractor orientation differed more from the memorized target orientation, we fit a von Mises to estimate the degree of variability in the report of the target (Figure 2B). There was a strong trend implicating that memory noise varied as a function of target-distractor difference ($F_{(8,56)} = 2.072$; $p = 0.054$), this trend was quadratic ($F_{(1,7)} = 5.214$; $p = 0.056$) implying noisier memory representations

when the target and distractor orientations differed more. Post-hoc *t*-tests showed no deviations from baseline (all $p > 0.14$).

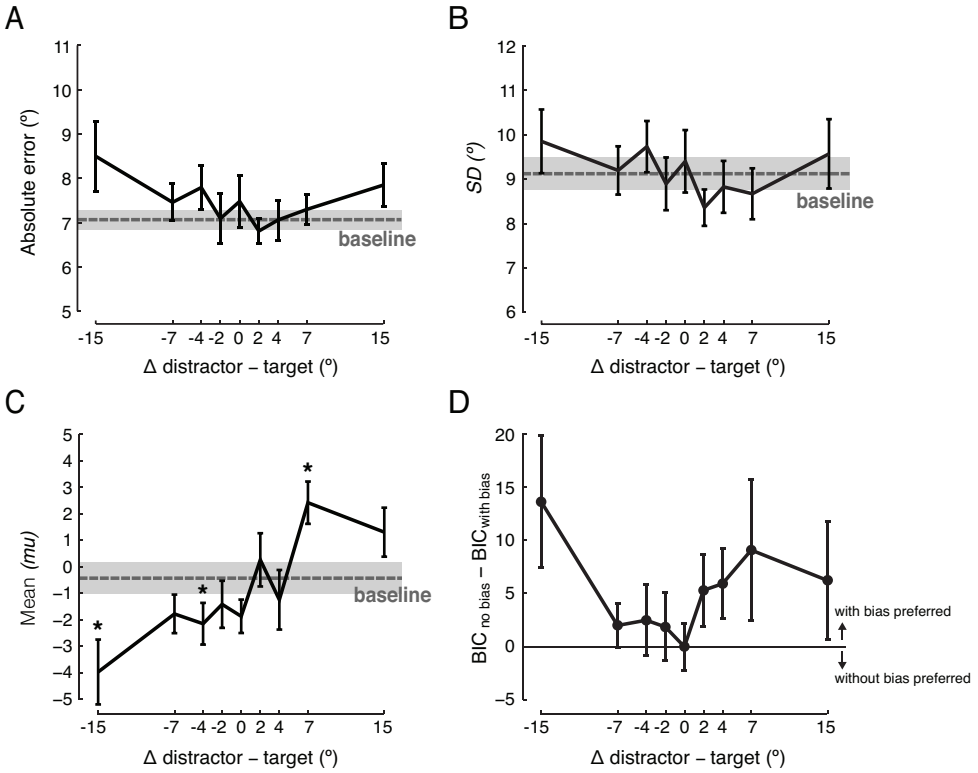


Figure 2. Results Experiment 1. **(A)** When the distractor orientation differs more from the remembered orientation (be it counter-clockwise or clockwise) the absolute error made by participants goes up relative to when the distractor orientation differs less from the remembered orientation. **(B)** The loss of accuracy at larger relative orientation differences between the target and distractor grating can be in part explained by a similar trend in memory precision (increase in *SD*). **(C)** The increase in the absolute error can furthermore be attributed to a shift in the entire response distribution *towards* the orientation of the distractor grating. This shift is more prominent on trials where the distractor grating differs more from the target, as indicated by a linear trend. **(D)** A von Mises with a bias term, modeling the distribution mean (*mu*) in addition to the distribution variability (*SD*), fits the data better than a von Mises without a bias term. This preference does not become significantly stronger at larger relative orientation differences between the target grating and the irrelevant grating. Group-averaged data is plotted in solid black lines with error bars representing ± 1 SEM; Grey dashed lines and shaded regions indicate the mean (± 1 SEM) on baseline trials during which no distractor was presented throughout the 3-second retention interval - obtained during separate experimental blocks. Asterisks indicate a significant difference from baseline trials.

In addition to a marginally noisier memory, larger orientation-errors arose due to a shift of the response-error distribution's mean (Figure 2C; $F_{(8,56)} = 12.404$; $p < 0.001$). This effect was linear ($F_{(1,7)} = 22.475$; $p = 0.002$), demonstrating that when the distractor orientation was rotated relative to the orientation in memory, participants' responses were shifted in the direction of the rotation. For a schematic overview of this 'attraction' effect, also see Figure 8 (General Discussion). Post-hoc *t*-tests demonstrated that the distribution mean was significantly different from the distribution mean during baseline trials when the target and distractor differed by –15°, –4°, and 7° ($p = 0.01$; $p = 0.013$; and $p = 0.007$ respectively).

We compared the fits from a von Mises with and without a parameter describing the distribution mean (Figure 2D). BIC difference values ('no bias' minus 'with bias') of > 0 indicated that a von Mises with bias term better fits the data than a von Mises without bias. Figure 2D shows that a von Mises with a bias term best described our data (i.e. testing ΔBIC against 0: $F_{(1,7)} = 5.722$; $p = 0.048$), though we did not find statistical support for the idea that it does so more for larger relative orientation distances between the target and distractor (i.e. testing whether ΔBIC differs for different target-distractor conditions: $F_{(8,56)} = 1.154$; $p = 0.343$).

Experiment 2

Methods

Participants. Participants in Experiment 2 were eight volunteers (7 female) aged 21 to 33 years old (M = 24.75; SE = 1.57), of whom four had already participated in Experiment 1. All participants had normal or corrected-to-normal vision and provided informed consent. The standing ethical committee of the Psychology and Neuroscience department at Maastricht University approved the study. Participants were naïve to the purpose of the

study and were reimbursed for their time (with the exception of two of the authors, RR and IB).

Stimuli and Procedure. Experiment 2 was virtually identical to Experiment 1, with two minor exceptions: First, the range of possible distractor orientations around the target was wider, spanning 90° in total. The distractor orientation could differ from the target orientation by –45°, –30°, –15°, –7°, 0°, 7°, 15°, 30°, or 45° and these conditions were presented in a randomly interleaved fashion. Secondly, instead of measuring a baseline in separate blocks as was done in Experiment 1, here we randomly interleaved trials without a distractor during retention. In total we collected 1000 trials divided over 5 blocks (~25 minutes per block).

Results and Discussion

Here we tested whether larger target-distractor differences would result in even larger shifts of the response distribution towards the distractor orientation, or might alternatively return to baseline, by expanding the range of relative orientation differences with respect to the range used in Experiment 1. Replicating our previous findings, Figure 3A shows that with the wider range of orientation differences, the absolute response error differed at various relative target-distractor orientation differences ($F_{(8,56)} = 3.899$; $p = 0.001$). As in Experiment 1, this effect was 'v-shaped', or quadratic ($F_{(1,7)} = 7.966$; $p = 0.026$), indicating bigger performance decrements when the relative orientation difference between target and distractor was larger, compared to when it was smaller. Post-hoc paired *t*-tests show that this is primarily due to the contrast between relatively small (i.e. 0° and 7° differences) compared to relatively large (i.e. 15°, 30°, and 45° differences) target-distractor differences (*p*-values between 0.001 and 0.308; $p = 0.09$ on average). Additionally, paired *t*-tests indicated that memory performance was negatively affected (compared to the no-distractor baseline) when the distractor was rotated –45° and –15° relative to the memory target ($p = 0.005$ and $p = 0.034$ respectively).

A von Mises was fit to examine the respective contributions of memory variability (Figure 3B) and mean response (Figure 3C). Memory noise differed across the various target-distractor conditions ($F_{(8,56)} = 3.872$; $p = 0.001$), and a quadratic trend indicated that memory became noisier as the target and distractor differed more, compared to when they differed less ($F_{(1,7)} = 6.539$; $p = 0.038$; for a schematic depiction of this 'v-shaped' effect, see Figure 8). Post-hoc paired *t*-tests show that contrasting relatively small (i.e. 0° and 7°) with relatively large (i.e. 15°, 30°, and 45°) target-distractor differences generally accounts for this 'v-shape' (*p*-values between 0.016 and 0.48; $p = 0.121$ on average). Finally, memory noise during trials with a distractor did not differ from no-distractor baseline trials ($F_{(1,7)} = 0.050$; $p = 0.829$), implying that the 'v-shaped' effect reflected a mixture of enhanced precision at smaller target-distractor differences, and reduced precision at larger target-distractor differences.

The mean response was shifted for various target-distractor difference conditions ($F_{(8,56)} = 10.589$; $p < 0.001$). This linear effect ($F_{(1,7)} = 20.421$; $p = 0.003$) indicated that, as in Experiment 1, the distractor orientation attracted the representation of a single orientation held in memory. The mean response was significantly shifted compared to the no-distractor baseline when the distractor was rotated –45°, –15°, –7°, 15°, 30°, and 45° relative to the memory target (paired *t*-tests $p = 0.008$; $p = 0.018$; $p = 0.006$; $p = 0.022$; $p = 0.035$ and $p = 0.008$ respectively).

Figure 3D shows that including a bias term to the von Mises did not result in significantly better data fits than not including such a bias term (i.e. testing ΔBIC against 0: $F_{(1,7)} = 3.146$; $p = 0.119$). Despite a shift in responses towards the distractor orientation (Figure 3C), adding a bias parameter did not help describe our data better at larger relative orientation differences. ($F_{(8,56)} = 0.890$; $p = 0.531$).

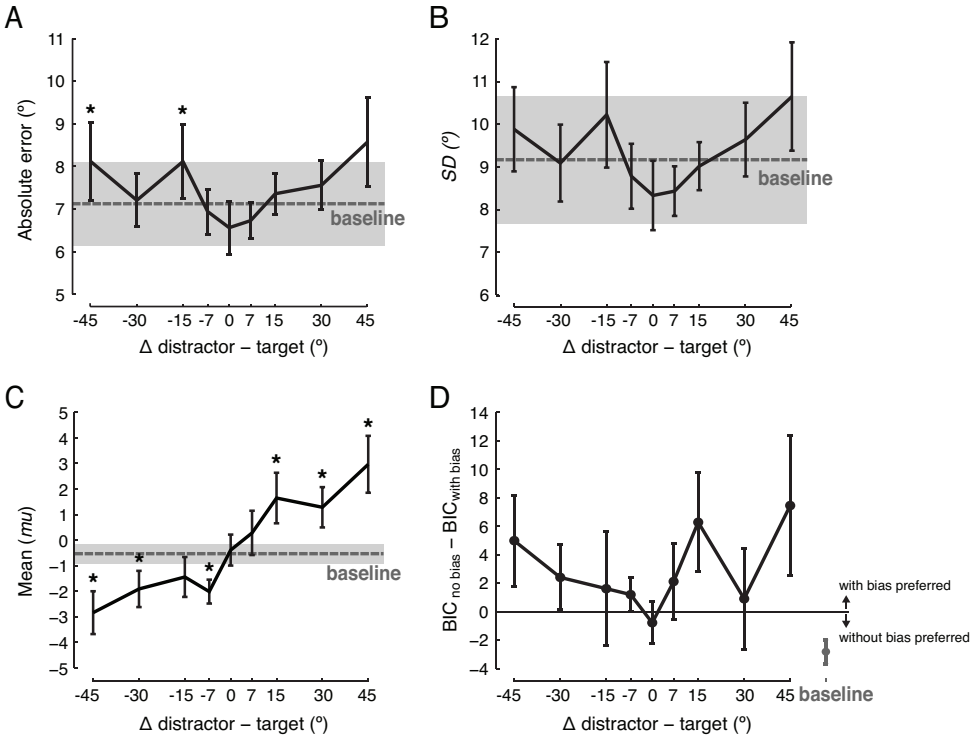


Figure 3. Results Experiment 2. **(A)** Participants make larger errors replicating an orientation in memory when the orientation of an irrelevant distractor grating (presented halfway through the retention interval) differed more from the memorized orientation. **(B)** Memory is noisier for the remembered orientation (larger *SD*) at larger relative target-distractor orientations, compared to less noisy memory (smaller *SD*) at smaller relative target-distractor orientations. **(C)** The error-response distribution shifts *towards* the orientation of the distractor grating; this shift is larger when difference in orientation difference between the two is larger. **(D)** Despite a shift in the mean of the response distribution, our data are no better described by a von Mises that includes this shift parameter, nor is there any evidence suggesting that a von Mises including a bias term fits the data better at larger relative orientation differences. Group-averaged data is plotted in solid black lines with error bars representing ± 1 SEM; Grey dashed lines and shaded regions indicate the no-distractor baseline mean ± 1 SEM derived from randomly interleaved trials. Asterisks indicate conditions for which there was a significant difference with the no-distractor baseline.

We did not find that the wider range of orientation differences presented here changed the attraction signature found in Experiment 1. Firstly, we did not observe that attraction returned to baseline at a 45° target-distractor difference. Secondly, there was no change in the magnitude of the distribution shifts between the first two experiments ($t_{(14)} = 0.329$; p

= 0.747): In Experiment 1 the distribution of response errors shifted from an average *mu* = –3.98° (when the distractor was rotated 15° clockwise relative to the memory target) to an average *mu* = 2.424° (when the distractor was rotated 7° relative to the memory target), resulting in a maximum observed shift of 6.4°. In Experiment 2 the maximum shift observed was 5.822° (from –2.852° to 2.971° in the –45° and 45° difference conditions respectively).

In addition, we found that memory noise did not differ between the two experiments (repeated measures ANOVA with a between-subject factor: $F_{(4,56)} = 0.843$; $p = 0.504$) when comparing the conditions both had in common (–15°, –7°, 0°, 7°, and 15° target-distractor differences). We furthermore calculated the maximum impact that the distractors had on precision (largest *SD* minus smallest *SD*), and found that the maximum impact in Experiment 1 (of 1.5° between target-distractor conditions of –15° and 2°) did not differ from the maximum impact in Experiment 2 (of 2.32° between target-distractor conditions of 0° and 45° difference) ($t_{(14)} = 0.654$; $p = 0.524$).

Experiment 3

Methods

Participants. Eight participants (5 female) took part in Experiment 3 (ages between 21 and 34 years, M = 26; SE = 1.64), including four volunteers who had previously participated in Experiments 1 and 2, and one who had previously participated in Experiment 2 only. Participants had normal or corrected-to-normal vision and provided their informed consent. The study was approved by the standing ethical committee of the Psychology and Neuroscience department at Maastricht University. With the exception of two of the authors (RR and IB), participants were reimbursed for their time and naïve to the purpose of the experiment.

Stimuli and Procedure. Stimuli and procedures in Experiment 3 (see Figure 4) were identical to those from Experiment 2, apart from three exceptions. We reduced the possible orientation differences between the two gratings to –45º, –15º, 0º, 15º, and 45º. Furthermore, participants were no longer instructed to ignore the second stimulus, but instead had to attend and remember it. The two stimuli had an equal probability of being cued once the test display was presented, with the cue being a black number ('1' or '2') displayed at fixation for as long as the test stimulus was on the screen. Thirdly, because both stimuli were always task relevant we did not obtain a baseline during which one of the two stimuli was not present. In total, participants performed 1000 trials divided over 5 blocks (~24 minutes each).

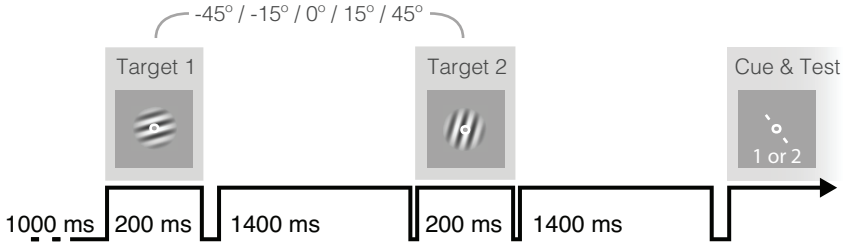


Figure 4. Trial sequence Experiment 3. Participants remembered two gratings, both presented for 200ms each. The first was randomly orientated and retained for 3 seconds. The second had an orientation that was rotated –45º, –15º, 0º, 15º, or 45º relative to the first grating (randomly interleaved) and was retained for 1400ms. A central number cue (not veridically depicted here in the interest of legibility) indicated which grating to report, and a test stimulus could be rotated by using a mouse to match the cued orientation in memory as precisely as possible. Participants clicked the mouse to continue to the next trial once satisfied with their response.

Results and Discussion

Here we investigated what would happen to the 'v-shaped' and 'attraction' effects found in the previous two experiments, when the second grating was task relevant. We also looked at how memory for the second orientation was affected by the first, and whether its representation would be prone to the same biases. Replicating our previous experiments,

we found that participants' responses were more erroneous as the difference between the first and second grating was larger (Figure 5A; $F_{(4,28)} = 12.881$; $p < 0.001$ with a quadratic trend $F_{(1,7)} = 43.567$; $p < 0.001$). This effect was due solely to the difference between conditions during which the target and distractor did not, and did share the same orientation (i.e. Δ non probed minus probed of 0° versus all other Δ non probed minus probed). Specifically, when the target and distractor differed by 45° the errors were larger than when the target and distractor differed by 0° (all paired *t*-test $p < 0.024$), and a similar trend was found comparing the conditions with 15° versus 0° differences (*p*-values between 0.011 and 0.066). None of the conditions where the target and distractor differed from one another revealed any differences in the absolute error (all $p > 0.284$). In other words, the 'v-shaped' effect levels off once the target and distractor start to differ. Finally, accuracy was better for the second grating compared to the first ($F_{(1,7)} = 6.05$; $p = 0.043$).

Memory variability (or *SD,* Figure 5B) mirrored the effects found in the absolute errors, indicating noisier memory when the two orientations differed more compared to when they differed less ($F_{(4,28)} = 10.547$; $p < 0.001$ with a quadratic trend $F_{(1,7)} = 44.893$; $p < 0.001$), and noisier memory for the orientation that was presented first ($F_{(1,7)} = 46.413$; $p < 0.001$). Post-hoc *t*-tests showed that when target-distractor differences existed (i.e. by 15° or 45°) memory was noisier than when target and distractor were of the same orientation (i.e. 0° difference) (all $p < 0.058$). Comparing all conditions where target-distractor differences existed yielded only one instance where the larger difference resulted in more memory variability than the smaller difference (for grating 2, comparing the –15° and 45° conditions: $p = 0.011$), while all other comparisons did not (all $p > 0.069$).

Memory for the first orientation was attracted towards the (now relevant) second orientation (solid black line in Figure 5C; $F_{(4,28)} = 2.79$; $p = 0.046$; linear trend $F_{(1,7)} = 5.711$; $p = 0.048$). However, the maximum distribution shift of 2.056° (defined as the largest clockwise shift minus the largest counterclockwise shift) was much reduced compared to Experiments 1 ($t_{(14)} = 3.4$; $p = 0.004$) and 2 ($t_{(14)} = 2.503$; $p = 0.025$), where the maximum shifts were 6.4° and 5.822° respectively. Distribution means for the second target (dashed

black lines in Figure 7C) show a different shift from those of the first target ($F_{(4,28)} = 4.647$; $p = 0.005$). In fact, there was a trend indicating that memory for the second target was shifted away from the orientation of the first target ($F_{(4,28)} = 1.808$; $p = 0.155$).

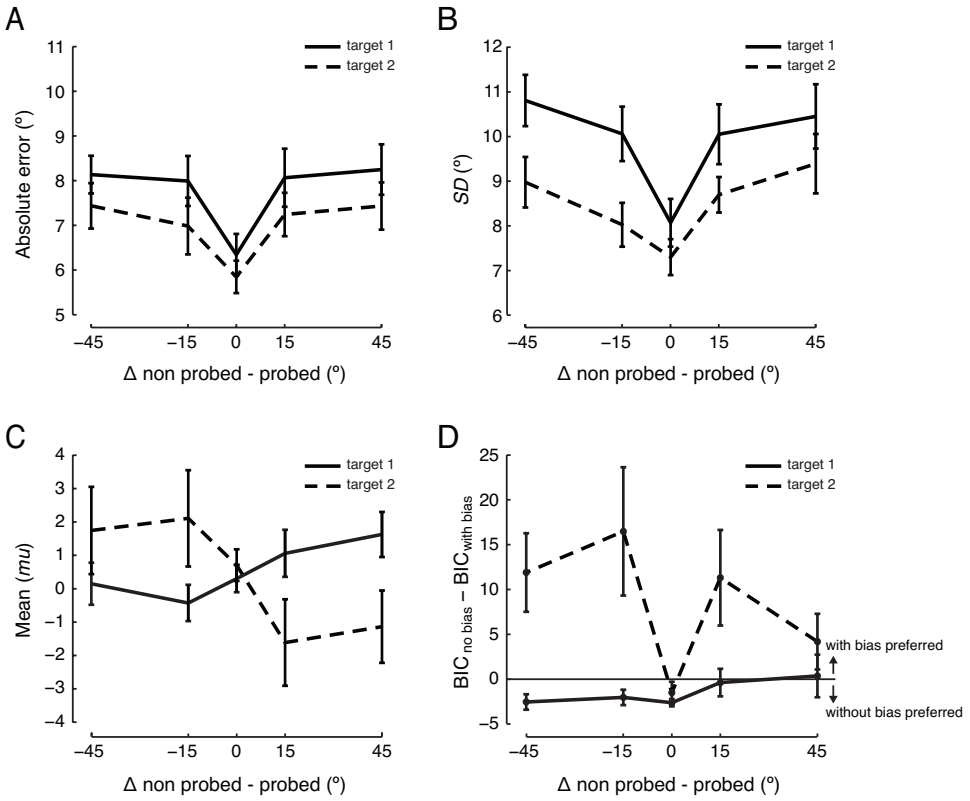


Figure 5. Results Experiment 3. **(A)** Replication errors were larger when two (remembered) orientations differed from one another. Moreover, performance was better for the orientation that was presented last. **(B)** The effects in (A) are mirrored by the variability of a von Mises that was fit to the distribution or error-responses. **(C)** The error distribution for responses to the first target shifts *towards* the orientation of the second target, but this attraction is smaller than the attraction in previous experiments where the second orientation was ignored. The error distribution for responses to the second target appears to be shifted *away* from the orientation of the first target, but this effect does not reach statistical significance. **(D)** Responses to target 1 are equally well described by a von Mises with and without bias term. Responses to target 2 are better fit when a bias term is included, and this benefit is more apparent when the two remembered orientations differed from one another. Group-averaged data for target 1 and 2 is plotted in solid black and dashed black lines respectively. Error bars represent ± 1 SEM.

In terms of which model best described the data the second and first target also differed (Figure 5D; $F_{(4,28)} = 4.575$; $p = 0.006$). The distribution of responses to the first target was equally well described by a von Mises with and without bias term (i.e. testing ΔBIC against 0: $F_{(1,7)} = 4.244$; $p = 0.078$) and this did not change across the various orientation differences between the two memory items ($F_{(4,28)} = 1.018$; $p = 0.415$). However, there was a clear benefit to include a bias term for the second target (i.e. testing ΔBIC against 0: $F_{(1,7)} = 5.200$; $p = 0.057$) which was not the same at the various relative orientation differences between the two memory items ($F_{(4,28)} = 4.936$; $p = 0.004$). Adding a bias parameter improved the fit for conditions during which the two targets differed in orientation, compered to when they did not differ (paired *t*-tests of Δ0° against –45°, –15°, 15°, and 45° difference conditions: $p = 0.016$; $p = 0.031$; $p = 0.036$; and $p = 0.069$ respectively).

Adding a bias parameter helped fit the data for the second target (Figure 5D), while a tentative distribution shift away from the first orientation (Figure 5C) did not reach significance. Exploring individual participant biases to the second target (Supplementary Figure 1) uncovered that the majority of participants (N = 5) showed 'repulsion' with reports shifted away from the orientation of the first target (by 9.13° on average; SE = 1.934), while the others (N = 3) showed an attraction (of 4.59° on average; SE = 1.133). Thus, all participants had a certain degree of bias in their responses to the second target, but the direction of that bias differed between individuals.

Why might performance be better for the second grating compared to the first? One possibility is that the task on the second grating was easier: once the first grating was presented and its orientation was known, the observer was inadvertently provided with information about the orientation of the second grating. Because the second grating could only have a limited number of orientations relative to the first, participants could have (implicitly) learned this relationship over time. However, this explanation is unlikely, as we found no evidence that memory for the second grating improved as participants were exposed to more experimental trials (Supplementary Figure 2). Alternatively, the benefit for the second target could have reflected the shorter retention duration over which it was

remembered. From the baseline data obtained as part of the first experiment we observed a 0.426° (SE = 0.218) accuracy reduction when information had to be remembered for 2 more seconds (comparing a 1 and 3-second retention interval). In this experiment the difference in retention intervals between the two targets was slightly shorter, (1.6 seconds), while the loss of accuracy was slightly larger (mean = 0.768°; SE = 0.312). While it's difficult to directly compare data from different participants in different experiments, it is conceivable that temporal decay played a role in the performance differences between the two targets. Finally, there is evidence from other studies suggesting that the last item of a series of sequentially presented – or fixated – stimuli is assigned more resources than previous items, making its representation in memory more precise in a way that cannot be explained by temporal decay (Bays & Husain, 2008; Gorgoraptis, Catalao, Bays, & Husain, 2011; Warden & Miller, 2007; Zokaei, Gorgoraptis, Bahrami, Bays, & Husain, 2011). Such an account could also explain the benefit enjoyed by the second grating in this experiment – the last in the series of two to-be-remembered stimuli.

Experiment 4

Methods

Participants. For Experiment 4 we tested ten participants (ages between 23 and 31 years, M = 24.3; SE = 0.79, 5 female). All ten participants were naïve to the purpose of the experiment, though one of them participated in all of the previous experiments reported here, and two others had previously participated in Experiment 3 only. Participants had normal or corrected-to-normal vision, provided informed consent, and all received monetary reimbursement for their time. The standing ethical committee of the Psychology and Neuroscience department at Maastricht University granted ethical approval for the experiment.

Stimuli and Procedure. Unique to Experiment 4 was that stimuli were viewed dichoptically through a mirror-stereoscope, presenting each of the two eyes with separate and independent information. Displaying stimuli dichoptically allowed us to capitalize on a phenomenon called binocular rivalry. This entails that two different images shown one to each eye will rival with each other, resulting in only one of the two reaching awareness while the other is suppressed.

Participants' eye-dominance was determined before the start of the experiment by a procedure that matches the relative strength of two rivaling images by adjusting their respective contrasts (adapted from: Pearson, Clifford, & Tong, 2008; Pearson, Rademaker, & Tong, 2011): Each run (of 100 trials in total) started with a 200-ms rivalry display, presenting a vertically oriented grating to the right eye, and a horizontally oriented grating to the left eye (grating stimuli were identical to others used throughout these experiments, unless mentioned otherwise). Participants indicated which grating (vertical or horizontal) they had perceived, after which they were adapted for 2 seconds to a full-contrast version of that pattern. Adaptation was used to facilitate a perceptual switch to the non-perceived image on the next trial, since the adapted pattern is less likely to reach awareness upon subsequent rivalry presentation (Pearson & Brascamp, 2008). When perception does not switch, this implies that the eye viewing the adaptor was the more dominant one. Initially, both monocular images were presented at 75% Michelson contrast, but if adaptation failed to induce a perceptual switch, the relative contrast of the two rivalry gratings was adjusted (by a factor of 2.5) on the subsequent trial: reducing stimulus contrast for the adapted (dominant) eye, and increasing it for the non-adapted (non-dominant) eye. This staircase procedure ensured that perception would eventually switch reliably between one pattern and the other. Once converged, the contrast values for the two eyes provided a measure of a participant's eye dominance, and its magnitude. Six out of the ten participants tested here were left-eye dominant.

In general, the outline of the task in this experiment (Figure 6A) was identical to that from Experiments 1 and 2, with any events occurring during the retention interval being

ignored. Stimuli were identical to those used in prior experiments with two exceptions: The first was that in addition to a second distractor grating (oriented –45°, –15°, 0°, 15°, or 45° relative to the target) also a circular 3°-diameter full-contrast checkerboard could be presented midway through the retention interval. The second exception was that, while the target was still presented at 20% Michelson contrast, the distractor was presented at 10% Michelson contrast (both $\pm$ 10% random uniform jitter) resulting in the distractor grating always having lower contrast than the target. This was done to increase the chance that the distractor would be suppressed when a checkerboard was simultaneously presented to the other eye. To further maximize the chances that the distractor grating would be rendered invisible when presented alongside a checkerboard, we stacked the odds in favor of the checkerboard by always presenting it to the dominant eye – where it was more likely to be consciously perceived. The distractor grating was ipso facto presented to the non-dominant eye where it was more likely to be suppressed. This also meant that the target grating was always presented to the non-dominant eye (as was the mouse probe).

Five types of events (Figure 6B), lasting 200 ms, could occur midway through the retention interval: (1) A distractor grating was presented to the non-dominant eye (same eye as target). (2) A distractor grating was presented to the dominant eye (different eye from target). (3) A distractor grating was presented to the non-dominant eye (same eye as target) while a checkerboard was simultaneously presented to the dominant eye (different eye from target). (4) A checkerboard was presented to the dominant eye (different eye from target). (5) No intervening stimulus was presented midway through the retention interval (true baseline).

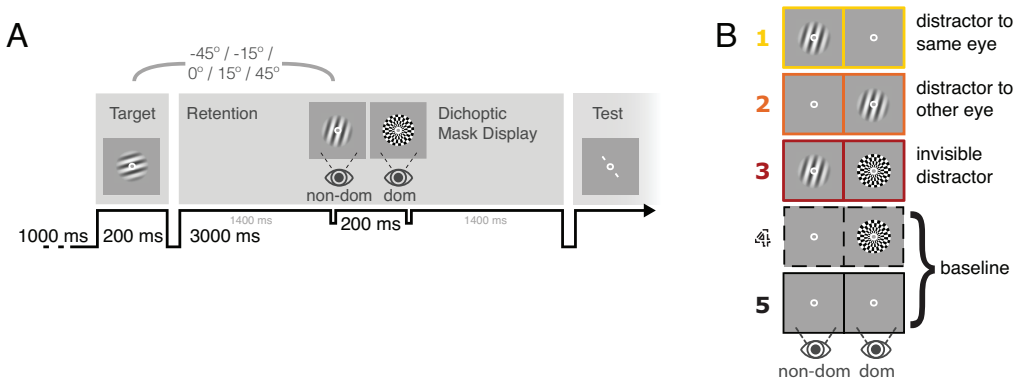


Figure 6. Trial sequence Experiment 4. **(A)** Participants remembered a randomly orientated grating presented to the non-dominant eye, ignored any events occurring during the retention phase, and replicated the orientation in memory 3 seconds later by rotating a test stimulus. **(B)** Five types of 200ms. events could occur during the retention phase: A distractor grating was visibly presented to the same (condition 1) or to the other (condition 2) eye, relative to the eye that the target was presented to. These first two conditions were perceptually identical. In condition 3, a distractor grating was presented to the same (non-dominant) eye and rendered invisible by a full-contrast checkerboard presented simultaneously to the other (dominant) eye (see also the example trial in A). Condition 4 is perceptually identical to condition 3 (participants only perceive the checkerboard), although only a checkerboard was presented to the dominant eye. Condition 5 was a true no-distractor baseline, with nothing being presented during the retention phase. The distractor gratings in conditions 1, 2, and 3 were rotated -45º, -15º, 0º, 15º, or 45º relative to the memorized grating (randomly interleaved). The color-coding corresponds to that in later figures for the purpose of convenience and comparability. The terms 'non-dom' and 'dom' are used to indicate viewing of stimuli with the non-dominant eye and the dominant eye respectively.

Note that conditions (1) and (2) were perceptually identical, since participants could not resolve which eye the stimuli originated from when viewing them through a mirror-stereoscope. These conditions thus replicated the perceptual experiences from participants in Experiments 1 and 2. Note that also conditions (3) and (4) were perceptually identical because the checkerboard was designed to always dominate perception, which meant that even if a distractor grating was presented simultaneously to the other eye (condition 3), only the checkerboard was consciously perceived. To ensure the latter was true, we included a 'visibility check' after all trials during which a checkerboard was presented. We asked participants to indicate on a 4-point scale what they had seen (1 = only the checkerboard; 2 = maybe something else; 3 = something else but unclear what; 4 =

something else and its orientation). Results from the visibility check showed that participants were indeed unable to perceive the distractor grating when presented dichoptically alongside the checkerboard. A rating of "4" was never given, and a rating of "3" was only given once throughout the entire experiment (this trial was removed and repeated later on during the run in which it had occurred). A rating of "2" was given on < 7% of trials (mean across participants 1.1% with SE = 0.613%), and this was independent of whether or not there was actually an irrelevant grating presented to the other eye ($t_{(9)}$ = 0.0007; p = 0.998). In total, participants performed 1700 trials divided over 10 blocks (~20 minutes each).

Results and Discussion

This experiment investigated the role of awareness: Is memory still biased by a distractor that people do not consciously perceive? A second question concerned binocular convergence: If biases persist when target and the distractor are presented to different eyes, interference occurs at a level of the visual hierarchy where information from the two eyes has been combined. Figure 7A shows that the 'v-shaped' effect for the absolute errors replicated when participants consciously perceived the distractor grating: We observed quadratic effects when the distractor was presented to the same (yellow line; $F_{(1,9)}$ = 12.451; p = 0.006) and the other (orange line; $F_{(1,9)}$ = 12.239; p = 0.007) eye relative to the eye presented with the memory target, and it made no difference to which eye the distractor was presented (compare yellow & orange lines; $F_{(1,9)}$ = 0.896; p = 0.368). Paired t-tests show that the quadratic effect arose from the comparison between the conditions where the target and distractor are the same (i.e. differ 0°) with all other conditions (i.e. target-distractor differences of 15° and 45°, all p < 0.057) and not from comparisons between conditions where target-distractor differences existed (all p > 0.12). Additionally, when a distractor was consciously perceived, performance suffered relative to the no-distractor baseline (same eye / yellow line: $F_{(1,9)}$ = 11.733; p = 0.008; other eye / orange line: $F_{(1,9)}$ = 19.117; p = 0.002), but only when the target and distractor differed in orientation (same

eye / yellow line: $F_{(4,36)} = 5.415$; $p = 0.002$; other eye / orange line: $F_{(4,36)} = 3.971$; $p = 0.009$; paired *t*-test *p*-values for Δ15° difference in other eye $p = 0.2537$; all other $p < 0.0345$). When the distractor grating was rendered invisible, no evidence of systematic changes in accuracy emerged across conditions (red line; $F_{(4,36)} = 0.450$; $p = 0.772$), and participants' responses were no different from when a checkerboard was presented by itself (compare red & dashed lines; $F_{(1,9)} = 0.045$; $p = 0.836$).

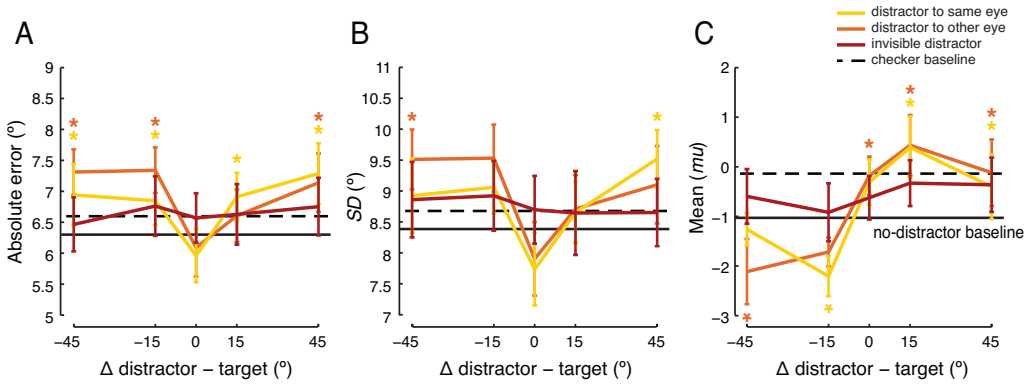


Figure 7. Results Experiment 4. **(A)** Participants make larger errors when a distractor orientation differs from a memorized orientation, but only when the distractor is consciously perceived, in which case it does not matter if it is presented to the same eye as the memory target or not **(B)** Memory is noisier when a consciously perceived but irrelevant distractor orientation differs from the orientation in memory. No such noise differences are observed when the distractor is rendered invisible by simultaneous presentation of a checkerboard stimulus to the other eye. **(C)** Only when a distractor orientation is consciously perceived (irrespective of the eye to which it is presented) does the response distribution shift *towards* the orientation of the distractor grating. Error bars represent ± 1 SEM. Asterisks indicate conditions for which there was a significant difference with the no-distractor baseline (colors correspond to eye-viewing conditions).

Estimates of the von Mises *SD* (Figure 7B) followed the absolute errors made by participants: Quadratic trends indicated that memory was noisier at larger target-distractor orientation differences, but only if the distractor grating was visibly presented to the same (yellow line; $F_{(1,9)} = 18.358$; $p = 0.002$) or the other eye (orange line; $F_{(1,9)} = 5.025$; $p = 0.052$) relative to the eye viewing the target. This was independent of eye-of-origin (compare yellow & orange lines; $F_{(1,9)} = 0.728$; $p = 0.416$). Again, the 'v-shape' or

quadratic effect could be accounted for by comparing 0° difference trials with 15° and 45° difference trials (paired *t*-test *p*-values between 0.001 and 0.075), and not because there were any changes between conditions where the target and distractor did not have the same orientation (all $p > 0.231$). Memory was noisier compared to the no-distractor baseline when the distractor was rotated 45° (and presented to the same eye: $t_{(9)} = 2.704$; $p = 0.024$) and when it was rotated –45° (and presented to the other eye: $t_{(9)} = 2.456$; $p = 0.036$). When the distractor was invisible, no systematic effects of memory variability were observed (red line; $F_{(4,36)} = 0.136$; $p = 0.968$), and memory variability was no different from trials on which a checkerboard was presented alone (compare red & dashed lines; $F_{(1,9)} = 0.078$; $p = 0.786$).

We replicated the shifts of the mean response when the distractor was visible (same eye / yellow line: $F_{(4,36)} = 9.97$; $p < 0.001$ and other eye / orange line: $F_{(4,36)} = 8.387$; $p < 0.001$), but not when the distractor was invisible (red line: $F_{(4,36)} = 0.914$; $p = 0.466$). The attraction towards visible distractors was independent of the eye to which they were presented (compare yellow & orange lines; $F_{(1,9)} = 0.016$; $p = 0.902$). Post-hoc *t*-tests demonstrated that the distribution mean was significantly different from the no-distractor baseline mean for conditions where the target and distractor differed by –15°, 15°, and 45° (same eye), and –45°, 0°, 15° and 45° (other eye) (all $p < 0.043$).

In terms of magnitude, the maximum distribution shifts were 2.598° and 2.547° for visible distractors presented to the same and other eye respectively. Compared to the previous experiments, this was similar in magnitude to the shift found in Experiment 3 (of 2.056°) where the second grating was task-relevant (same eye: $t_{(16)} = 0.636$; $p = 0.534$; other eye: $t_{(16)} = 0.681$; $p = 0.505$), whereas it was smaller than shift-magnitudes found in Experiments 1 and 2 (of 6.4° and 5.822° respectively) where the second grating was irrelevant (all $p < 0.031$). During this experiment participants were expected to ignore all events occurring during the retention interval, but it is likely that some deployment of attention occurred nonetheless because participants were effectively performing a second task in parallel to the primary memory task. Namely, they had to conclude trials where a

checkerboard was presented with a perceptual report (the visibility check – see Method), which required a certain degree of attention towards stimulus events taking place during retention. This could explain why the shift-magnitudes found here more closely resembled those from Experiment 3 (where the second grating was attended), than to those from Experiments 1 and 2 (where the second grating was ignored).

General Discussion

The experiments described here revealed systematic biases for a single orientation in memory that arose in response to interfering information (for a schematic overview, see Figure 8). First of all, when a second (to-be-ignored) distractor grating was presented during retention – its orientation parametrically varied relative to the remembered orientation – larger differences between the memorized and distractor orientations led to noisier memory compared to smaller differences between the two orientations. Additionally, the orientation represented in memory was attracted towards the distractor orientation, biasing the correct answer by ~3°. Increasing the range of target-distractor differences did not affect the magnitude of these two biases, while other factors such as attention and awareness did have an influence. Memory attraction was reduced when the interfering information was made task relevant, and memory biases were completely abolished when participants did not consciously perceive the distractor.

Memory interference by means of attraction has been fairly well established (Dubé et al., 2014; Huang & Sekuler, 2010a; Nemes et al., 2011; 2012). Here we revealed changes in memory noise due to a distractor orientation presented during retention, accounting for an additional source of memory error that has not been previously demonstrated. Do these noise changes affect memory for better of for worse? When plotting memory noise (*SD*) against the target-distractor differences examined throughout our four experiments, we consistently found a quadratic effect. This 'v-shaped' effect could comprise of increased memory precision when target-distractor differences were small (a change for the better), decreased memory precision when target-distractor differences were large (a

change for the worse), or a mixture of the two. Predominantly, memory precision on trials where a distractor was present (of any orientation) did not differ from precision during a no-distractor baseline, implying a mixture of enhanced and impaired precision at smaller and larger target-distractor differences respectively. Furthermore, changes in memory noise leveled off once target-distractor differences became sufficiently large. For example, no additional changes were observed between target-distractor differences of 15° and 45°, where the degree of memory noise neither increased nor returned to baseline.

Increasing the range of orientation differences between the target and distractor within a single experiment did not impact the magnitude of attraction and noise biases to which memory is susceptible (Experiment 2). This refutes a perceptual averaging account (Dubé et al., 2014; Huang & Sekuler, 2010a) whereby larger target-distractor differences should have resulted in a stronger attraction. Attraction was also not found to return to baseline, which should necessarily occur once a target and distractor differ by 90° – no direction can be inferred when the two are orthogonal. Previous studies did observe that attraction returned to baseline once target-distractor differences fell outside of the bandwidth often assumed of early sensory processing (Nemes et al., 2011; 2012; Van der Stigchel, Merten, Meeter, & Theeuwes, 2007). For orientation, early sensory tuning bandwidths are estimates around 40°~52° (Albright, 1984; De Valois, Yund, & Hepler, 1982), implying that our target-distractor range with a maximum of 45° might have been just shy of revealing such a return to baseline. That said, there was no indication of this in Experiment 2 (Figure 3C). Instead we found that both biases in attraction, as well as precision, scaled to the range of target-distractor differences presented within a single experiment.

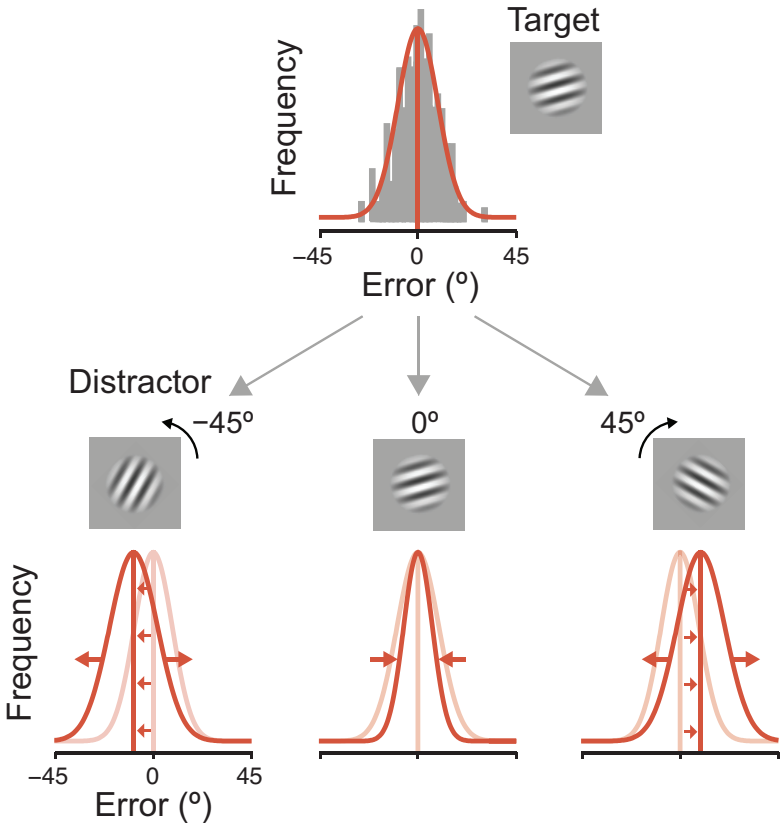


Figure 8. General overview of the main findings. When a single orientation is maintained in memory (top row), a second orientation presented during retention (middle row of distractor gratings) affects memory performance because (i) memory is noisier when the target and distractor differ, compared to when the target and distractor are the same. This is illustrated by changes in the width of the error distribution in response to the target (solid orange/grey in the top row, lighter orange/grey in the bottom row) showing a mixture of narrowing when target and distractor are the same (bottom row, middle column, solid orange/grey) and broadening when target and distractor differ (bottom row, outer two columns, solid orange/grey). Memory performance is furthermore affected because (ii) the orientation represented in memory is attracted towards the distractor orientation. This is illustrated by the distribution shift (bottom row, outer two columns, solid orange/grey) in the direction of the distractor orientation by ~3°. For illustrative purposes we have schematized and exaggerated the effects on memory noise and distribution means, and only present –45° and 45° difference conditions.

When the interfering information was made task relevant (Experiment 3), the extent to which the first orientation was attracted towards the second was much reduced compared

to when information from the second grating could be completely discarded (as in Experiments 1 and 2). Additionally, we found that memory for orientation showed attraction only when the first (but not the second) of two task relevant orientations was probed. These findings directly contradict a comparable study where memory for spatial frequency was biased strongly towards a non-probed spatial frequency irrespective of which of the two gratings was probed, and this attraction was reduced when the non-probed item could be ignored (Huang & Sekuler, 2010a). Thus, an attentional manipulation rendering a distractor either task relevant or irrelevant yielded markedly contradictory findings depending on whether memory for orientation or for spatial frequency was tested. Memory for these two features might be somehow fundamentally different, or subtle dissimilarities in methodology could be at fault. Regardless, it's clear that a perceptual averaging account of memory interference (Dubé et al., 2014; Huang & Sekuler, 2010a) did not hold for our data. For one, if the degree to which two stimuli are averaged depends on their respective weights, which are in turn determined by mechanisms of selective attention, stronger attraction towards an ignored distractor is unlikely (Experiments 1 & 2 versus Experiment 3). Furthermore, perceptual averaging can only explain memory attraction, which is inconsistent with the majority of participants in Experiment 3 demonstrating a substantial repulsion.

Repulsion was expressed by the second (task relevant) orientation being rotated in a direction opposite from the first. One possible explanation is that the second grating was already perceived as repelled away from the first and thus remembered as such (Scocchia, Cicchini, & Triesch, 2013). Another way to look at repulsion is by realizing that when two successive stimuli have different orientations there could be an implied rotation, or 'movement'. For example, if the second grating was clockwise relative to the first, a 'clockwise rotation' was implied. Memory for moving targets is biased in accordance with expectations people have about the physical characteristics of real-life moving objects, such as momentum, gravity, or friction (Hubbard, 1995a; 1995b). In our example this means a shift of both representations in the direction of the implied clockwise rotation: For the first grating this would look like 'attraction' towards the second (clockwise)

grating. For the second grating the rotation would be clockwise relative to its own actual orientation, making it look like a 'repulsion' relative to the first grating. Nonetheless, the fact that a subset of participants showed attraction rather than repulsion under the same circumstances indicates that individuals used different strategies, and that the direction of the bias exerted by a second stimulus is not fixed. To some extent, memory biases could be decisional rather then perceptual in nature (Park, Rademaker, & Tong, 2014), which contradicts theories assuming interference occurring within the visual hierarchy (Huang & Sekuler, 2010a; Magnussen, 2000).

Memory biases were abolished when participants did not consciously perceive the interfering distractor orientation, implying that awareness of the distractor was necessary for it to influence representations held in memory (Experiment 4). This directly contradicts earlier work where a subliminal distractor negatively affected memory accuracy (Silvanto & Soto, 2012). The absence of distractor interactions during rivalrous stimulus presentation furthermore suggests a locus of interference at a relatively late stage of visual processing – while suppressed information is still observed in V1 (Maier et al., 2008), this signal peters out along the visual hierarchy (Blake & Logothetis, 2002). When the distractor orientation was consciously perceived, the eye to which the distractor orientation was presented (relative to the target) did not impact short-term memory biases, implying a brain locus beyond the point of binocular convergence (Experiment 4). The memory biases observed in the present study are unlikely to arise at early sensory levels, which is supported by the existence of binocular cells in V1 (Parker, 2007; Poggio, Motter, Squatrito, & Trotter, 1985) in combination with our finding that interference occurred on binocularly combined signals that required awareness in a rivalrous setting.

One alternative explanation for the biases reported here might be that on some portion of trials participants reported the wrong orientation. As the memorized and distractor gratings differed more, such mistakes would lead to increases in fitted noise and larger apparent distribution shifts towards the distractor orientation. This is unlikely the case; first of all our task was trivially easy, making it improbable that errors due to misreporting

would occur frequently, if at all. The error distributions (Supplementary Figure 3) confirm this. Secondly, errors due to misreporting would amplify memory biases as the range of target-distractor differences increased, which was not found to be the case. Finally, we directly compared data fits from a von Mises to fits from a bimodal model that described the data in terms of a mean, circular variance, and probability of responses to the wrong orientation. We only included conditions where the target and distractor had different orientations, and where the second grating was visibly presented. A von Mises fit our data better in all experiments (Experiment 1: $F_{(1,7)} = 897.384$; $p < 0.001$; Experiment 2: $F_{(1,7)} = 21.30$; $p = 0.002$; Experiment 3: $F_{(1,7)} = 11.714$; $p = 0.011$; Experiment 4: $F_{(1,9)} = 3.68$; $p = 0.087$), further ruling out the possibility of faulty reports by our participants.

Despite explicit instructions to ignore the second grating, its orientation nevertheless affected participants' responses to the target orientation. This can be explained neither by weighted averaging and magnet metaphors, nor by misreporting of the stimulus. An alternative way of interpreting our findings is in the context of a Bayesian framework, which explains perception by integrating sensory information (or likelihood) with expectations about the world (prior). Within this framework, perceptual biases can emerge when people integrate prior information about a stimulus in a Bayesian fashion, following certain mathematical rules. Prior information can be derived from repeated long-term exposure to natural image statistics, such as the over-representation of horizontal and vertical information in natural scenes (Girshick, Landy, & Simoncelli, 2011) or default illumination and observer viewpoints being from above (Gerardin, Kourtzi, & Mamassian, 2010; Mamassian & Goutcher, 2001; Mamassian & Landy, 1998). Such priors might be implemented as early as primary sensory cortex, interacting with sensory evidence at very early stages of visual processing (Kok, Brouwer, van Gerven, & de Lange, 2013; Vintch & Gardner, 2014). However, priors are not necessarily static, and can be generated or updated in an experimental setting within a relatively short time (Brady, Konkle, & Alvarez, 2009; Chopin & Mamassian, 2012; Körding & Wolpert, 2004; Turk-Browne, Scholl, Chun, & Johnson, 2009).

In our experiment, the second grating had a limited number of possible orientations relative to the first, which meant it retroactively provided participants with information about the remembered orientation. Once learned, such statistical information could have been used as a prior and integrated it into participants' responses. When trying to infer the memorized orientation, statistical information introduced by a distractor could have been more or less equally informative irrespective of the range of all possible distractors spanned around the target orientation: While a larger range of distractor orientations around the target can be seen as a chance for stronger biases, it also implies a larger variance on the prior, and these two factors might cancel out. Potentially then, such integration could explain why we found that memory biases scaled with different target-distractor ranges in Experiments 1 and 2. However, we did not find support for the idea that repeated exposures to the distractor orientations in our study resulted in statistical learning: When we calculated memory attraction on a trial-by-trial basis, distractor effects were observed immediately, and did not evolve over time (Supplementary Figure 4).

Perhaps the integration of irrelevant feature information into short-term memory is an unavoidable feature of the system (Marshall & Bays, 2013). Such involuntary integration could deplete memory resources and reduce precision, which could explain our finding of larger variability when target and distractor orientations differed more. Alternatively, integration of information can be viewed as a common strategy, used because in many contexts it actually improves the accuracy of report (Huttenlocher, Hedges, & Duncan, 1991; Huttenlocher, Hedges, & Vevea, 2000). Memory interference could also be related to serial dependence in human vision, which assumes that both past and present inputs are used to inform perception. For example, a memorized spatial frequency (Huang & Sekuler, 2010a; Huang & Sekuler, 2010b), location (Simmering, Schutte, & Spencer, 2008) or perceived orientation (Whitney & Fischer, 2014) is attracted towards information from the recent past. Memory contents could be affected in a similar manner, with a distractor serving as a past influence on a current memory representation.

There is evidence that also other contexts are able to systematically bias the contents of visual short-term memory. Ensemble statistics are an example, such as a reported bias towards the average size of items in memory (Alvarez, 2011; Brady & Alvarez, 2011), or a central tendency biasing memory towards the average or prototypical value of a stimulus set (Dubé et al., 2014; Freyd & Johnson, 1987; Huang & Sekuler, 2010a; Huttenlocher et al., 1991; Spencer & Hund, 2002; Wilken & Ma, 2004). Eye movements can bias memory in the direction of the saccade (Bays & Husain, 2008). A spatial reference frame can bias memory for location away from salient axes (Simmering et al., 2008; Spencer & Hund, 2002). Notably, memory for spatial location was found to shift towards the location of a task irrelevant distractor (Van der Stigchel et al., 2007), which parallels our findings for orientation. This shift only occurred when the distractor was relatively close by, which is reminiscent of the idea that attraction only occurs within a limited bandwidth (Nemes et al., 2011; 2012; Sneve et al., 2015). When memory uncertainty is higher (for example if items have to be retained for longer durations, if contrast is low, if set size is large, etc.), people are thought to rely more strongly on such biases to postulate a response (Dubé et al., 2014; Vintch & Gardner, 2014).

Thus, many types of interactions can arise amongst memory representations, or between stored information and incoming perceptual input. The extent of these interactions depends strongly on the metric distance between stimulus features relative to each other. These so-called metric interactions offer another compelling interpretation for the memory distortions reported here. To explore this in more detail we borrowed from work on dynamic neural network models consisting of a layered, neutrally plausible architecture. This class of model can capture a wide variety of behaviors by assuming the existence of perceptual, as well as short- or long-term memory model-layers that interact via local excitatory and lateral inhibitory interactions (Johnson, Spencer, Luck, & Schöner, 2009; Simmering et al., 2008; Wei, Wang, & Wang, 2012). Such connectivity achieves sustained activation during working memory, but also leads to metric interactions between memory items presented at different times (Simmering et al., 2008) different spatial locations (Wei et al., 2012), and between stored information and perceptual input

(Johnson et al., 2009). While items are maintained in working memory in these models, processes such as merging (Wei et al., 2012) and biasing (Simmering et al., 2008) can occur among representations in memory. A tentative explanation of the biases observed in our study could be derived from combining these delay-period dynamics with consequences from newly arriving sensory input – shown to add a subtle peak of activation in a neural model's memory layer (Johnson et al., 2009). If a small peak of distractor-centered excitation merges with the maintained representation one would expect a memory shift toward the distractor, as well as an increase in variance (Wei et al., 2012).

This raises the question which brain mechanisms implement interactions between memory and sensory processes, and give rise to memory biases observed in the literature. A recent neuroimaging study demonstrated that when people remembered a single spatial frequency, and were presented with a distractor whose spatial frequency differed by 1 cycle per degree (but not 2 cycles per degree), activity to the distractor was suppressed in visual areas V3 and V4 (Sneve et al., 2015). This aligns with predictions of a 'Mexican hat' shaped interaction profile from neural network models (Johnson et al., 2009; Simmering et al., 2008). Critically, modulations in V4 influenced activity in earlier areas V1-V3, where a stronger suppression correlated with performance decrements. This finding also complements previous psychophysics (Magnussen, 2000) by alluding to a possible brain mechanism involving interference and top-down influences from V4 initiating behaviorally relevant changes in V1.

However, this study only revealed how memorized information affected processing of a distractor, without addressing what happened to the memory representation itself. A lot is already known about how high-level cognitive states, such as memory, can alter the readiness or modulate responses to new sensory inputs through feedback activity, biasing sensory neurons and perception (Johnson et al., 2009; Lui & Pasternak, 2011; Mendoza, Schneiderman, Kaul, & Martinez-Trujillo, 2011; Mendoza-Halliday, Torres, & Martinez-Trujillo, 2014; Miller & Desimone, 1994; Miller, Li, & Desimone, 1991; 1993; Scocchia et

al., 2013; Zaksas & Pasternak, 2006). Indeed, working memory contents can guide selective attention (Downing, 2000; Soto, Hodsoll, Rotshtein, & Humphreys, 2008), and increase cortical excitability (Cattaneo, Pisoni, Papagno, & Silvanto, 2011). Our findings clearly demonstrate that dynamic interactions between bottom-up sensory information and top-down cognitive states (such as the maintenance of visual memories) work both ways, with newly perceived sensory information also biasing the processing of information that is actively maintained in memory. However, the mechanisms by which perception might bias representations in visual short-term memory are much less understood.

While it remains elusive whether memory interference serves a functional purpose, studying interference provides a window into the computational processes supporting visual memory by exploring how such mechanisms inoculate memories against the myriad of new images entering the eyes in rapid succession. Against a backdrop of complementary imaging, physiology, and psychophysical work, the current findings help further our understanding of the dynamic interactions involved in the maintenance of visual memories, and inform us about the fundamental question of how memories are stored.

References

Albright, T. D. (1984). Direction and orientation selectivity of neurons in visual area MT of the macaque. *Journal of Neurophysiology*,*52*(6), 1106–1130.

Alvarez, G. A. (2011). Representing multiple objects as anensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131.

Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Bays, P. M., Catalao, R. F. G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11.

Bennett, P. J., & Cortese, F. (1996). Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. *Vision Research*, *36*(2), 233–238.

Berens, P. (2009). CircStat: a MATLAB toolbox for circular statistics. *Journal of Statistical Software*, *31*(10), 1–21.

Blake, R., & Logothetis, N. K. (2002). Visual competition. *Nature Reviews Neuroscience*, *3*(1), 13–21.

Bosch, S. E., Jehee, J. F. M., Fernández, G., & Doeller, C. F. (2014). Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *The Journal of Neuroscience*, *34*(22), 7493–7500.

Brady, T. F., & Alvarez, G. A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

Brady, T. F., Konkle, T., & Alvarez, G. A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology General*, *138*(4), 487–502.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Cattaneo, Z., Pisoni, A., Papagno, C., & Silvanto, J. (2011). Modulation of Visual Cortical Excitability by Working Memory: Effect of Luminance Contrast of Mental Imagery. *Frontiers in Psychology*, *2*, 1–9.

Chopin, A., & Mamassian, P. (2012). Predictive properties of visual adaptation. *Current Biology*, *22*(7), 622–626.

De Valois, R. L., Yund, E. W., & Hepler, N. (1982). The orientation and direction selectivity of cells in macaque visual cortex. *Vision Research*, *22*(5), 531–544.

Deutsch, D. (1970). Tones and Numbers: Specificity of Interference in Immediate Memory. *Science*, *168*(3939), 1604–1605.

Deutsch, D. (1973). Interference in memory between tones adjacent in the musical scale. *Journal of Experimental Psychology*, *100*(2), 228–231.

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6), 467–473.

Dubé, C., Zhou, F., Kahana, M. J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research, 96*, 8–16.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433.

Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229.

Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 259–268.

Fukuda, K., Awh, E., & Vogel, E. K. (2010). Discrete capacity limits in visual working memory. *Current Opinion in Neurobiology*, *20*(2), 177–182.

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 128–134.

Gerardin, P., Kourtzi, Z., & Mamassian, P. (2010). Prior knowledge of illumination for 3D perception in the human brain. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(37), 16309–16314.

Girshick, A. R., Landy, M. S., & Simoncelli, E. P. (2011). Cardinal rules: visual orientation perception reflects knowledge of environmental statistics. *Nature Neuroscience*, *14*(7), 926–932.

Gorgoraptis, N., Catalao, R. F. G., Bays, P. M., & Husain, M. (2011). Dynamic updating of working memory resources for visual objects. *The Journal of Neuroscience*, *31*(23), 8502–8511.

Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Huang, J., & Sekuler, R. (2010a). Distortions in recall from visual memory: two classes of attractors at work. *Journal of Vision*, *10*(2): 24, 1–27.

Huang, J., & Sekuler, R. (2010b). Attention protects the fidelity of visual memory: Behavioral and electrophysiological evidence. *The Journal of Neuroscience, 30*(40), 13461–13471.

Hubbard, T. L. (1995a). Cognitive representation of motion: evidence for friction and gravity analogues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 241–254.

Hubbard, T. L. (1995b). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin & Review*, *2*(3), 322–338.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376.

Huttenlocher, J., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology General*, *129*(2), 220–241.

Johnson, J. S., Spencer, J. P., Luck, S. J., & Schöner, G. (2009). A Dynamic Neural Field Model of Visual Working Memory and Change Detection. *Psychological Science*, *20*(5), 568–577.

Kok, P., Brouwer, G. J., van Gerven, M. A. J., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *The Journal of Neuroscience 33*(41), 16275–16284.

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244–247.

Lalonde, J., & Chaudhuri, A. (2002). Task-dependent transfer of perceptual to memory representations during delayed spatial frequency discrimination. *Vision Research*, *42*(14), 1759–1769.

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Lui, L. L., & Pasternak, T. (2011). Representation of comparison signals in cortical area MT during a delayed direction discrimination task. *Journal of Neurophysiology*, *106*(3), 1260–1273.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature*, *17*(3), 347–356.

Magnussen, S. (2000). Low-level memory processes in vision. *Trends in Neurosciences*, *23*(6), 247–251.

Magnussen, S., & Greenlee, M. (1992). Retention and disruption of motion information in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(1), 151–156.

Magnussen, S., & Greenlee, M. (1999). The psychophysics of perceptual memory. *Psychological Research*, *62*(2), 81–92.

Magnussen, S., Greenlee, M. W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, *31*(7-8), 1213–1219.

Maier, A., Wilke, M., Aura, C., Zhu, C., Ye, F., & Leopold, D. (2008). Divergence of fMRI and neural signals in V1 during perceptual suppression in the awake monkey. *Nature Neuroscience*, *11*(10), 1193–1200.

Mamassian, P., & Goutcher, R. (2001). Prior knowledge on the illumination position. *Cognition*, *81*(1), B1–9.

Mamassian, P., & Landy, M. S. (1998). Observer biases in the 3D interpretation of line drawings. *Vision Research*, *38*(18), 2817–2832.

Marshall, L., & Bays, P. M. (2013). Obligatory encoding of task-irrelevant features depletes working memory resources. *Journal of Vision*, *13*(2):21, 1–13.

McKeefry, D. J., Burton, M. P., & Vakrou, C., (2007). Speed selectivity in visual short term memory for motion. *Vision Research, 47*, 2418–2425.

Mendoza, D., Schneiderman, M., Kaul, C., & Martinez-Trujillo, J. (2011). Combined effects of feature-based working memory and feature-based attention on the perception of visual motion direction. *Journal of Vision*, *11*(1):11, 1–15.

Mendoza-Halliday, D., Torres, S., & Martinez-Trujillo, J. C. (2014). Sharp emergence of feature-selective sustained activity along the dorsal visual pathway. *Nature Neuroscience*, *17*(9), 1255–1262.

Miller, E. K., & Desimone, R. (1994). Parallel neuronal mechanisms for short-term memory. *Science*, *263*(5146), 520–522.

Miller, E. K., Li, L., & Desimone, R. (1991). A neural mechanism for working and recognition memory in inferior temporal cortex. *Science*, *254*(5036), 1377–1379.

Miller, E. K., Li, L., & Desimone, R. (1993). Activity of neurons in anterior inferior temporal cortex during a short-term memory task. *The Journal of Neuroscience*, *13*(4), 1460–1478.

Nemes, V. A., Parry, N. R., Whitaker, D., & McKeefry, D. J. (2012). The retention and disruption of color information in human short-term visual memory. *Journal of Vision*, *12*(1):26, 1–14.

Nemes, V. A., Whitaker, D., Heron, J., & McKeefry, D. J. (2011). Multiple spatial frequency channels in human visual perceptual memory. *Vision Research*, *51*(23), 2331–2339.

Nilsson, T. H., & Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 141–150.

Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, *16*(2), 332–350.

Park, Y., Rademaker, R. L., & Tong, F. (2014). Both variations in perceptual sensitivity and decisional response bias contribute to visual working memory performance. *Journal of Vision*, *14*.

Parker, A. J. (2007). Binocular depth perception and the cerebral cortex. *Nature Reviews Neuroscience*, *8*(5), 379–391.

Pasternak, T., & Zaksas, D. (2003). Stimulus specificity and temporal dynamics of working memory for visual motion. *Journal of Neurophysiology*, *90*, 2757–2762.

Pearson, J., & Brascamp, J. (2008). Sensory memory for ambiguous vision. *Trends in Cognitive Sciences*, *12*(9), 334–341.

Pearson, J., Clifford, C. W. G., & Tong, F. (2008). The Functional Impact of Mental Imagery on Conscious Perception. *Current Biology*, *18*(13), 982–986.

Pearson, J., Rademaker, R. L., & Tong, F. (2011). Evaluating the Mind's Eye: The Metacognition of Visual Imagery. *Psychological Science*, *22*(12), 1535–1542.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Phillips, W. A. (1974). On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics*, *16*(2), 283–290.

Poggio, G. F., Motter, B. C., Squatrito, S., & Trotter, Y. (1985). Responses of neurons in visual cortex (V1 and V2) of the alert macaque to dynamic random-dot stereograms. *Vision Research*, *25*(3), 397–406.

Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, *6*(2), 461–464.

Scocchia, L., Cicchini, G. M., & Triesch, J. (2013). What's‘ up’? Working memory contents can bias orientation processing. *Vision Research*, *78*, 46–55.

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-Specific Delay Activity in Human Primary Visual Cortex. *Psychological Science*, *20*(2), 207–214.

Silvanto, J., & Soto, D. (2012). Causal evidence for subliminal percept-to-memory interference in early visual cortex. *NeuroImage*, *59*(1), 840–845.

Simmering, V. R., Schutte, A. R., & Spencer, J. P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research*, *1202*, 68–86.

Sneve, M. H., Alnæs, D., Endestad, T., Greenlee, M. W., & Magnussen, S. (2012). Visual short-term memory: activity supporting encoding and maintenance in retinotopic visual cortex. *NeuroImage*, *63*(1), 166–178.

Sneve, M. H., Sreenivasan, K. K., Alnæs, D., Endestad, T., & Magnussen, S. (2015). Short-term retention of visual information: Evidence in support of feature-based attention as an underlying mechanism. *Neuropsychologia*, *66*(C), 1–9.

Soto, D., Hodsoll, J., Rotshtein, P., & Humphreys, G. W. (2008). Automatic guidance of attention from working memory. *Trends in Cognitive Sciences*, *12*(9), 342–348.

Spencer, J. P., & Hund, A. M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology General*, *131*(1), 16–37.

Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision*, *13*(10):9, 1–8.

Turk-Browne, N. B., Scholl, B. J., Chun, M. M., & Johnson, M. K. (2009). Neural evidence of statistical learning: Efficient detection of visual regularities without awareness. *Journal of Cognitive Neuroscience*, *21*(10), 1934–1945.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785.

Van der Stigchel, S., Merten, H., Meeter, M., & Theeuwes, J. (2007). The effects of a task-irrelevant visual event on spatial working memory. *Psychonomic Bulletin & Review*, *14*(6), 1066–1071.

Vintch, B., & Gardner, J. L. (2014). Cortical correlates of human motion perception biases. The Journal of Neuroscience, *34*(7), 2592–2604.

Warden, M. R., & Miller, E. K. (2007). The Representation of Multiple Objects in Prefrontal Neuronal Delay Activity. *Cerebral Cortex (1991)*, *17*, i41–i50.

Wei, Z., Wang, X. J., & Wang, D. H. (2012). From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization. *The Journal of Neuroscience, 32*(33), 11228–11240.

Whitney, D., & Fischer, J. (2014). Serial dependence in visual perception. *Nature Neuroscience*, 1–9.

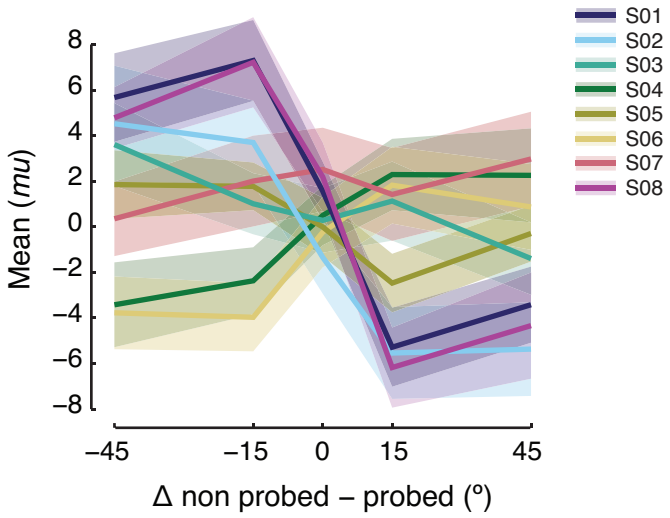
Wilken, P., & Ma, W. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120-1135.

Zaksas, D., & Pasternak, T. (2006). Directional Signals in the Prefrontal Cortex and in Area MT during a Working Memory for Visual Motion Task. *The Journal of Neuroscience*, *26*(45), 11726–11742.

Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.

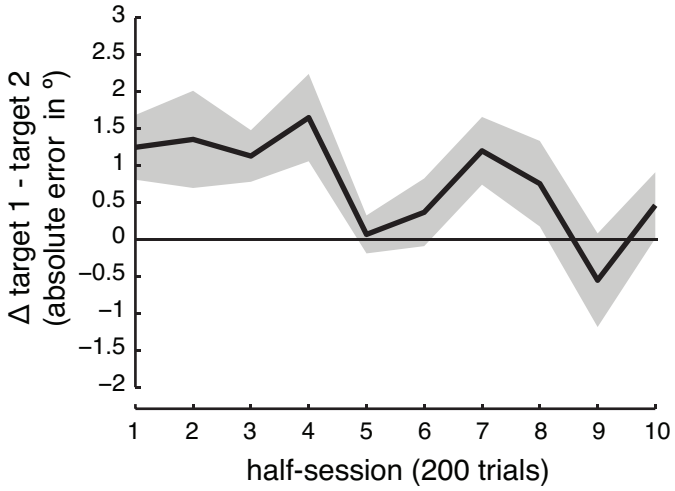
Zokaei, N., Gorgoraptis, N., Bahrami, B., Bays, P. M., & Husain, M. (2011). Precision of working memory for visual motion sequences and transparent motion surfaces. *Journal of Vision*, *11*(14): 2, 1–18.

Supplementary Figure 1



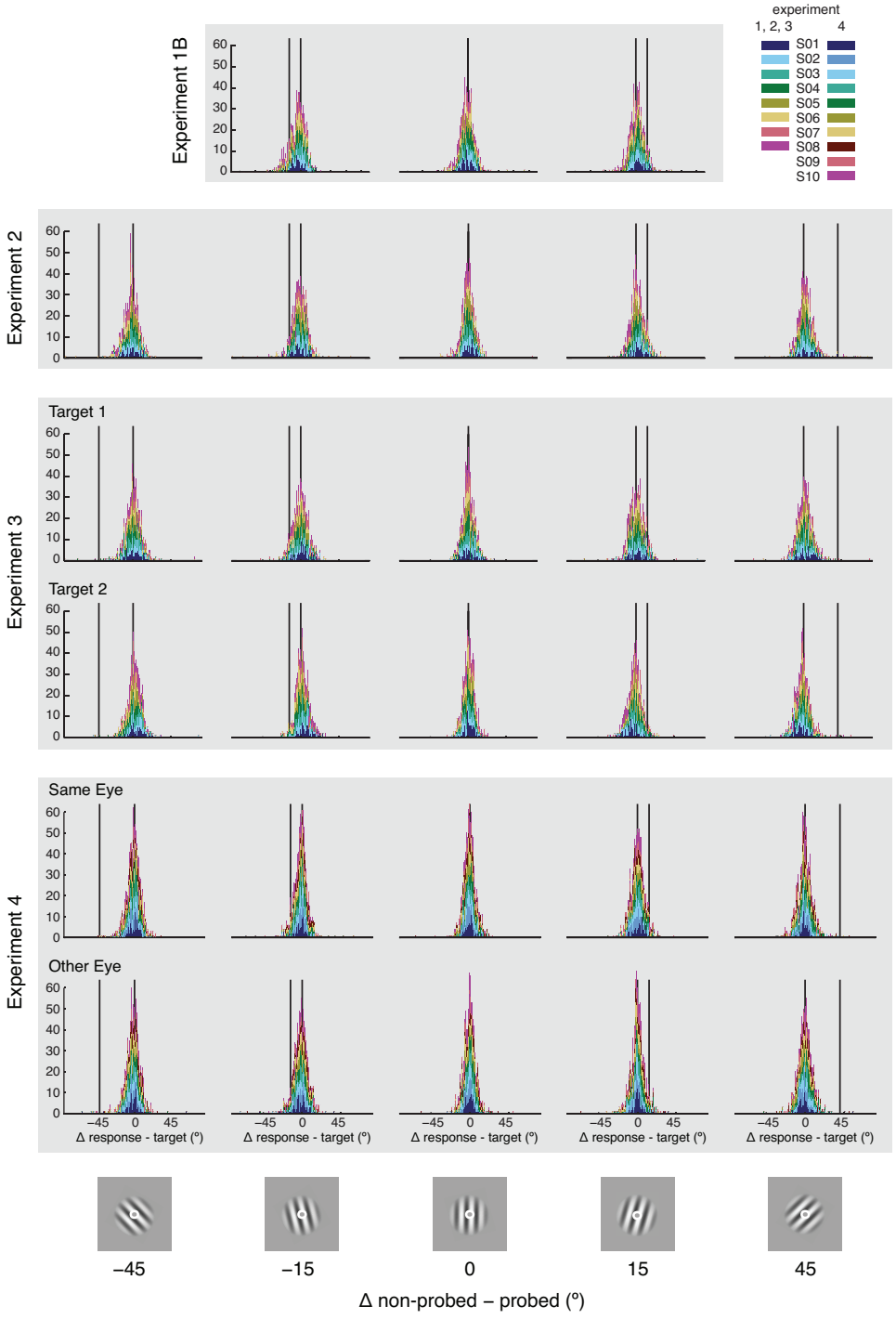
Supplementary Figure 1. When reporting the second of two to-be-remembered orientations, different participants show different biases relative to the orientation of the item that was memorized first. For participants 1, 2, 3 and 8 memory of the second orientation is shifted in a direction opposite to the first orientation. For participants 4 and 6 the reverse happens, and reports of the second orientation are attracted towards the first orientation. For participants 5 and 7 there is no significant bias in either direction. Here we used a Bayesian approach to fit individual subject data (described in more detail in Suchow et al., 2013), which constructs a full probability distribution over the model parameters, and uses a non-informative prior. This allows us to obtain 95% confidence intervals for individual participants (shaded regions) around the Maximum of the posterior distribution (MAP estimate) that serves as a point estimate for each parameter value (solid colors).

Supplementary Figure 2



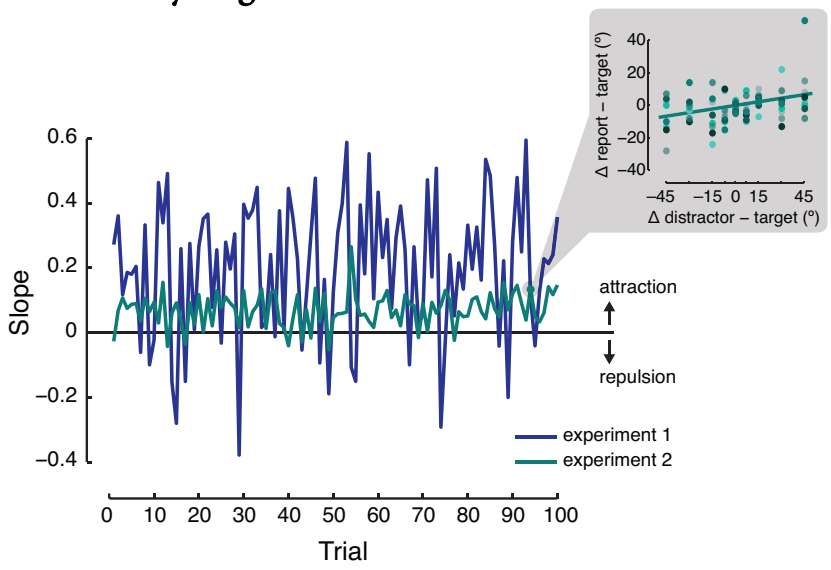
Supplementary Figure 2. Relative performance for the two memory targets over time. As soon as the first grating is shown, there is information available about the possible orientation of the second grating because the second grating can only have a limited number of orientations relative to the first. For example, if the first grating happens to be oriented at 36º (a truly random value between 1º and 180º), the second grating can only be 171º, 21º, 36º, 52º, or 81º. Note that, except for the two authors, participants had no idea about the systematic relationship between the two orientations, although learning (explicit or implicit) might have taken place during the thousand experimental trials participants were exposed to. We examined whether this might explain the overall performance benefit for the second grating compared to the first. For each 200 trials (half of a block) we calculated how much better performance for the second grating was compared to the first (absolute error target 1 – absolute error target 2), where values > 0 indicate an advantage for the second grating. The figure shows that even if participants learned the relationship between the two gratings over time, this did not increase the advantage enjoyed by the second grating. Instead, performance becomes more similar for the two gratings as participants progress through the experiment, with the initial performance benefit for the second target dwindling over time ($F_{(1,7)} = 6.507$; $p = 0.038$). Thus, it is unlikely that memory for the second grating was better due to participants having knowledge about its possible orientation prior to its presentation. The black line plots the group-averaged data with shaded grey regions representing $\pm$ 1 SEM.

Supplementary Figure 3



← **Supplementary Figure 3.** Error histograms for all experiments (selected conditions) and all subjects. The memory biases (in the mean response and memory noise) found throughout our study do not originate from participants mistakenly reporting the wrong orientation. Each histogram plots response frequency against the response errors (reported orientation - target orientation). Grey panels represent the different experiments in this study, showing histograms for Experiments 1B, 2, 3 (both targets), and 4 (only visibly presented irrelevant gratings where memory biases were detected). Per grey panel, the histograms display the distribution of responses for conditions where the relative orientation differences were -45º, -15º, 0º, 15º, and 45º. Only these relative orientation differences are shown here to ease comparison between experiments, and for the sake of conciseness. Different colors represent individual participants, while the black vertical lines in each histogram represent the target orientation (at 0º) and the orientation of the second grating presented in that particular condition.

Supplementary Figure 4



Supplementary Figure 4. *We performed a trial-by-trial analysis to observe how shifts of the distribution mean evolved over time. For every n^{th} trial of a given target-distractor condition we calculated the 'bias' on that trial (or simply put: the difference between target and response), and paired this with the corresponding n^{th} trials 'biases' from the full range of target-distractor differences. We plotted these 'biases' for all subjects against target-distractor conditions (see insert, different colors represent different participants), and determined the slope of a line fitted to these points. The slope is taken as a measure of the response shift on the n^{th} trials. The main panel depicts how the response shift evolves over time for Experiments 1 (dark blue) and 2 (green-blue) and demonstrates that memory was shifted (i.e. biased towards the irrelevant distractor) from the beginning onwards, and this shift did not increase over the course of the experimental trials.*

Chapter 3

Modeling false memory for orientation under the influence of irrelevant distractors

Manuscript in preparation:

Rademaker R.L., Mamassian, P., Sack A.T.

Modeling false memory for orientation under the influence of irrelevant distractors.

Abstract

Short-term memory for simple visual features is susceptible to interference: when a memory target and distractor differ from one another on the task-relevant feature, the representation in memory is attracted towards that of the distractor. Recent work confirmed such attraction for orientation memory, showing a bias in the mean response towards an irrelevant orientation. Here we aim to elucidate some of the computational principles that might underlie these false memories for orientation. We model data from two experiments during which observers remembered one random target orientation over a three second delay, and adjusted a probe to match their memory afterwards. Halfway through the delay an irrelevant orientation was presented that differed parametrically with respect to the target. The attraction found in these experiments was well described by a weighted averaging model, which assumed that responses were a simple weighted average of the two orientations. A second model, assuming a decisional stage as well as the multiplicative integration of target and distractor distributions, was also able to adequately predict the attraction bias. Finally, both models suggested that the distractor's influence was smaller when the range of orientation differences between the target and distractor was larger.

Introduction

Our visual world, while seemingly detailed and rich, is one of limited spatial and temporal resolution. Short-term maintenance of visual information is of critical importance for the construction of a rich percept from such sparse inputs, and for keeping relevant information online to realize cognitive goals. The study of visual short-term memory aims to explore the mechanisms by which information is stored, and how robust storage can be achieved in the presence of an unrelenting stream of new inputs arriving from the eyes, carrying a strong potential for interference.

Indeed, for the process of perception to be achieved in an adequate manner, it is pivotal that a visual buffer stores veridical representations. However, it's been shown that visual short-term memory is susceptible to systematic biases, and memory contents can be significantly distorted to favor new and interfering inputs (Huang & Sekuler, 2010). Such memory interference has been uncovered during trivially easy tasks, when people remembered only one simple feature, such as a single color (Nemes, Parry, Whitaker, & McKeefry, 2012), spatial frequency (Bennet & Cortese, 1996; Dubé, Zhou, Kahana, & Sekuler, 2014; Magnussen, Greenlee, Asplund, & Dyrnes, 1991; Nemes, Whitaker, Heron, & McKeefry, 2011), velocity (Magnussen & Greenlee, 1992; McKeefry, Burton, & Vakrou, 2007), location (Van der Stigchel, Merten, Meeter, & Theeuwes, 2007), or orientation (Rademaker, Bloem, De Weerd, & Sack, 2015). Specifically, these studies employed a procedure known as 'memory masking': With a single target item in memory, a second task-irrelevant item (called a 'mask' or 'distractor') is presented during the delay. The distractor can match the memory target (for example, a blue distractor presented while remembering a blue target), or differ from the memory target (for example, a green distractor presented while remembering a blue target). When the target and distractor differed along the task relevant dimension (but not a task irrelevant dimension, see Magnussen et al., 1991), discrimination thresholds were increased by about twofold (Magnussen et al., 1991; Magnussen & Greenlee 1992; McKeefry, 2007), due to a shift in

the mean response towards the irrelevant distractor (Huang & Sekuler, 2010; Nemes et al., 2011; Nemes et al., 2012; Rademaker et al., 2015).

Thus, non-veridical memories arise due to an attraction towards task-irrelevant information. False memories of this nature might be quantified as a simple weighted combination of the target and distractor, by virtue of the bias towards an intermediate value (Huang & Sekuler, 2010; Dubé et al., 2014). Indeed, one study investigated memory biases for spatial frequency by cuing one of two consecutively presented gratings either before or after their presentation (Huang & Sekuler, 2010). Responses were always biased towards the spatial frequency of the non-cued stimulus, although the attraction was smaller with a pre-cue than with a post-cue. It was concluded that behavioral responses consisted of a weighted average of the two spatial frequencies, and that selective attention acted on these weights to curtail the influence of a distractor when it was pre-cued and could be ignored (Huang & Sekuler, 2010; Dubé et al., 2014).

Recently, a series of four experiments investigated memory for orientation in a 'memory masking' context, whereby a single mnemonic item was replicated using a mouse probe. The first two experiments parametrically varied the orientation of an irrelevant distractor, presented during the delay, relative to a randomly oriented memory target. The range of target-distractor differences spanned 30° and 90° for the first and second experiment respectively (see also Figure 1). While target-distractor differences could be much larger in the second experiment, both experiments uncovered an attraction towards the distractor of approximately equal magnitude. This finding was incompatible with weighted averaging, as averaging would predict a stronger attraction when a target and distractor differ more. The interfering second stimulus was made task relevant in a third experiment, reducing the attraction it exerted on the first stimulus. This *weaker* attraction towards an attended orientation contrasted strongly with previous findings for spatial frequency, where attraction was *stronger* when an item was attended, and called into question the role of attention on averaging weights (Huang & Sekuler, 2010; Dubé et al., 2014). Furthermore, the majority of participants in the third experiment by Rademaker et al.

(2015) reported the second orientation as further away from the first than it actually was. This repulsion discouraged an interpretation involving weighted averaging still more, as averaging dictates that only attraction can exist. Finally, all four experiments showed noisier memories at larger target-distractor differences compared to smaller target-distractor differences, which weighted averaging also cannot account for.

These findings for orientation memory show that the current way in which memory interference is quantified needs to be revised. We applied two quantitative models to data from the first two experiments by Rademaker et al. (2015), and attempted a first pass at an alternative explanation for biases in the mean report. Ultimately, our aim will be to also explain the observed changes in variance, as well as changes with attentional state. As a first benchmark model we looked at weighted averaging, fitting target and distractor weights in order to predict responses. Given the range invariance of the attraction effect, we expected that when the target-distractor range increased from experiment 1 to experiment 2, the distractor weights would consequently be reduced. We also evaluated a variance dependent version of this model, in which the weights were estimated based on stimulus reliability. Secondly, a decisional component has been implicated in memory for orientation (Park, Rademaker, & Tong, 2014; Rademaker et al., 2015). Our second model explored this further by assuming a decisional stage during which observers decided where an orientation value lay relative to the cardinal axes. Target and distractor representations were modeled as tuned orientation distributions, assuming larger variance for the distractor and multiplicative integration –model fits reflected the width of the distractor distribution relative to that of the target.

Methods

Participants. Eight volunteers participated in Experiment 1 (ages between 21 and 29; 6 female), and eight volunteers participated in Experiment 2 (ages between 21 and 33; 7 female) of whom four also took part in the first experiment. All participants had normal or corrected-to-normal vision, and provided informed consent. The study took place

under the approval of the standing ethical committee of the Psychology and Neuroscience department at Maastricht University. Participants received monetary reimbursement for their time, and were naïve to the purpose of the study (with exception of two participants involved in running the experiment, including author RR).

Stimuli and Procedure. Participants viewed all stimuli in a dark room on a luminance-calibrated CRT monitor with 1280 x 1024 resolution and 60 Hz refresh rate. Visual stimuli were generated using MATLAB 7.5.0 (R2007b) and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997) on a PC running windows XP. Participants were seated at a viewing distance of 57 cm, and were instructed to maintain fixation throughout, aided by a chinrest and a white central fixation bull's eye (0.5° diameter). Stimuli consisted of centrally presented oriented gratings (3° diameter; spatial frequency 2 c/°; 20% Michelson contrast $\pm$ 10% uniform jitter; phase randomized). Gratings were presented within a wide Gaussian envelope (*sd* = 2.5°) on a uniform grey background sharing the same mean luminance (40.8 cd/m). Responses were obtained with a mouse-probe comprising of the white bull's eye and an interrupted white line (3° gap) that could rotate around fixation, each segment of the line was 0.125° long and 0.025° wide.

The task in both experiments was to remember, as precisely as possible, a single randomly oriented (1–180°) target grating presented for 200ms (Figure 1A). After a 3s retention interval the mouse-probe was presented at an initially random orientation, also between 1 and 180°, and by moving the mouse participants were able to rotate this dial to match the orientation in memory. Once satisfied with the response, a mouse click triggered the onset of a 1s inter-trial interval, after which the next trial commenced. There were no time constraints during the response phase, and precision was emphasized.

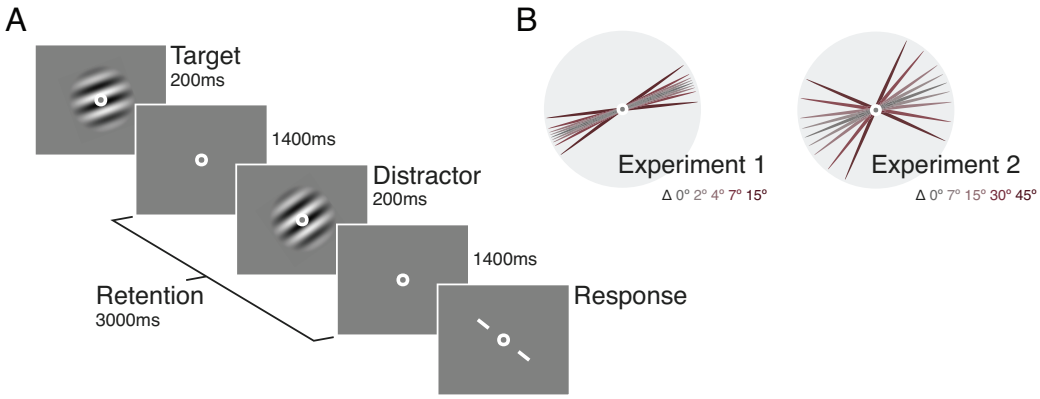


Figure 1. Trial sequence and relative distractor orientations **(A)** Participants viewed a randomly oriented target orientation, and remembered it as precisely as possible for a duration of 3 seconds, after which they replicated the memorized orientation by means of a mouse probe rotating around fixation. Midway through the delay an irrelevant distractor orientation was presented which was supposed to be ignored. The interval between successive trials was one second. **(B)** The orientation of the distractor was defined relative to that of the target. In Experiment 1 the orientation could either be the same as that of the target (Δ 0º condition) or rotated by 2º, 4º, 7º, or 15º relative to the target orientation (see left panel). In Experiment 2 the distractor orientation could also be the same (Δ 0º) or was rotated by 7º, 15º, 30º, or 45º relative to the target orientation (right panel). The various Δ target-distractor conditions were randomly interleaved within experimental blocks. The example trial in (A) shows the Δ –15º condition, with the distractor rotated 15º counter clockwise relative to the target.

During the retention interval a second, irrelevant grating was presented for 200ms. Participants were instructed to ignore this distractor. The orientation of the distractor was parametrically varied with respect to the orientation of the target. For Experiment 1 distractor orientations were sampled within a limited range, and more densely around the target (Gaussian δ=25), resulting in distractors that were rotated by –15º, –7º, –4º, –2º, 0º, 2º, 4º, 7º or 15º relative to the orientation of the target (Figure 1B, left panel). Thus, Experiment 1 consisted of nine Δ target-distractor conditions, with negative and positive Δ indicating counter clockwise and clockwise rotations of the distractor relative to the target respectively (i.e. condition names reflect the distractorº – targetº). Experiment 2 also consisted of nine Δ target-distractor conditions, but the range of possible differences was expanded such that the distractor could differ from the target by –45º, –30º, –15º, –7º, 0º, 7º, 15º, 30º, or 45º (Figure 1B, right panel). All Δ target-distractor conditions were

presented in a randomly interleaved fashion. A baseline during which no distractor was presented was obtained either in separate blocks (Experiment 1) or randomly interleaved with distractor trials (Experiment 2). After a brief practice, 100 trials per condition were collected for each participant.

Analyses. Data analysis were performed in MATLAB using custom functions and the Circular Statistics Toolbox (Berens, 2009). The circular mean (μ) was calculated to indicate central tendencies of data, and the circular variance (σ) to indicate data variability. Models were fit using a non-linear fitting procedure ('nlinfit' in MATLAB).

Results

Memory for a single orientation was tested in the presence of an irrelevant distractor orientation that could differ from the target by nine possible Δ. The range of Δ differed between Experiments 1 and 2.

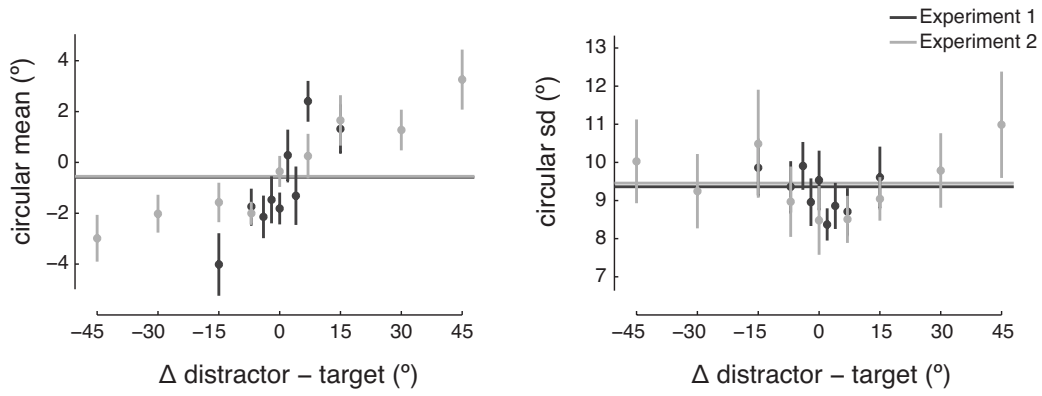


Figure 2. Mean and variance of target responses during Experiments 1 and 2. The left panel plots the circular mean against the various Δ target-distractor conditions. For both experiments the mean response was shifted in the direction of the to-be-ignored distractor. The right panel plots the circular variance against the various Δ target-distractor conditions. For both experiments the variance was larger when the target and distractor differed more, compared with when they were more similar.

When calculating the circular mean and standard deviation (as in Rademaker et al., 2015) for all Δ target-distractor conditions we observed an attraction of the mean target response towards the distractor (Figure 2, left panel), and increased variability when the target and distractor differed more compared with when they differed less (Figure 2, right panel).

Emergence over time. Comparing the two experiments showed that the range of Δ target-distractor orientations tested within each experiment did not affect the magnitude of the observed memory biases (in both mean and variance). In other words, the impact of a distractor further away in orientation space was not stronger (Rademaker et al., 2015). This raised the possibility that target-distractor relationships might be learned over time. If so, this knowledge could be integrated as prior knowledge when postulating a response. To investigate whether knowledge about the range of Δ target-distractor conditions was learned over time, we performed two time-resolved analyses. Here, we calculated the circular mean for each of the Δ target-distractor conditions, over a sliding window of 5 consecutive experimental trials. Figure 3, only plotting the most extreme Δ target-distractor conditions for illustrative purposes, shows that biases were present from the get go. A disadvantage of the sliding window analysis was that it required 5 trials to obtain the earliest bias estimate in each of the nine Δ target-distractor conditions, and consequently a total of 45 trials across conditions. Learning might occur on a smaller time scale. Nevertheless, this analysis corroborates the second, previously performed analysis (Rademaker, et al., 2015; Supplementary Figure 4). This analysis estimated the bias in report within the first 9 trials, by plotting the circular mean against the Δ target-distractor conditions (as in Figure 2) and estimating the slope. Positive slopes were observed from the earliest time point onward. Thus, false memories for orientation, showing a bias towards an irrelevant distractor, appear to be immediate and do not emerge over time.

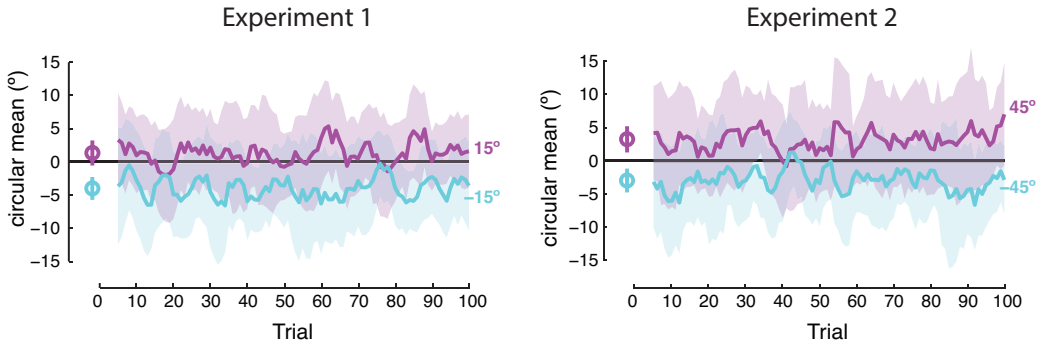


Figure 3. Emergence of bias in the mean response over time. The circular mean is plotted against trial number, with each data point representing the mean over the interval [trial–4 trial]. For both experiments only the largest Δ target-distractor conditions are shown for illustrative purposes. Response biases in the direction of the distractor could be observed from the earliest response mean estimate.

Default biases. What might bring about these non-veridical memories? A first step to consider before answering this question is to investigate performance in the absence of a distractor. There are good reasons to believe that all orientations should not be considered equal, for example, perception is more sensitive for orientations at or near the vertical and horizontal axes compared with all others, a phenomenon known as the 'oblique effect' (Appelle, 1972). To examine this and other possibilities, we looked at 'default' biases in memory for orientation in the absence of a distractor. Figure 4 shows data from the no-distractor baseline trials for Experiment 1 (left panels) and 2 (right panels) collapsed across all participants. The top panels plot raw response errors against target orientation, showing that the errors were not uniformly distributed around the correct answer (an error of 0°). Two performance measures (circular standard deviation and mean) were extracted from the raw errors, using a 20° sliding window analysis along all possible target orientations (1–180°). Middle panels show that (especially for Experiment 2) the standard deviation was smaller around cardinal orientations, indicating an oblique effect. Moreover, the bottom panels show that participants had a strong tendency to report slightly off-cardinal orientations as shifted further away from cardinal than they actually were (bottom panels). Thus, a clear 'default' bias was observed for the mean report, implying repulsion away from the cardinal axes.

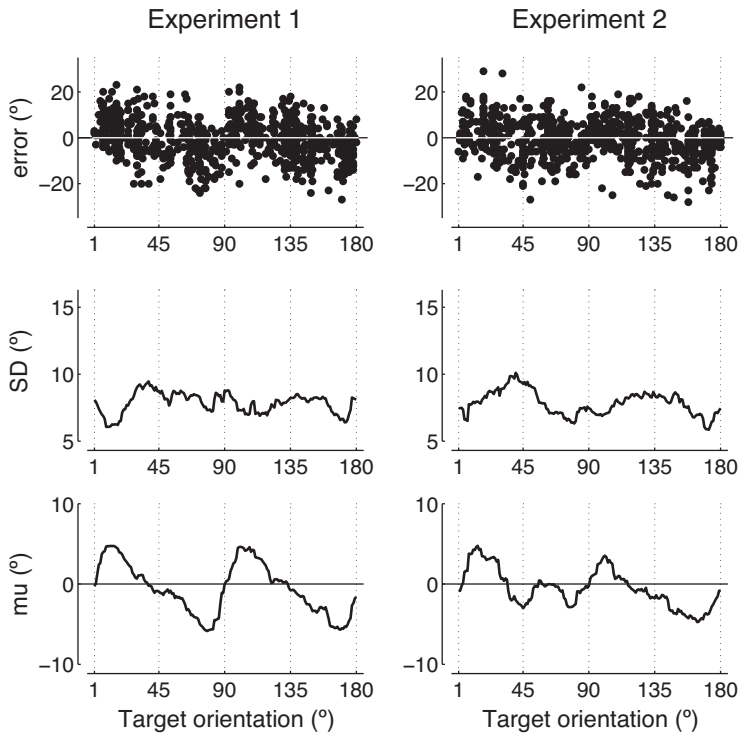


Figure 4. Baseline errors and parameters. Top panels show response errors (in ° and without errors > 3 SD), middle panels show the circular SD (in °), and bottom panels show the circular mean – all plotted against the target orientation. Circular SD and μ were calculated over a 20° sliding window (of the errors shown in the top panels). Thus, for each target orientation we calculated these two parameters based on responses to that target orientation ± 10 degrees. Such a windowed approach results in some smoothing of the data (with larger windows resulting in more smoothing).

Cardinal axes. Default biases in the mean report (Figure 4, bottom panels) demonstrated a clear influence of cardinal axes on the responses in an orientation replication task. These natural boundaries imposed on orientation-space might serve as a spatial reference frame, and could factor into an explanation of false memories under the influence of a distractor. To explore how memory response biases were affected by the horizontal and vertical meridians we separated all experimental trials by whether or not the target and distractor crossed a cardinal axis. For example, a trial on which a 4° target was presented with a distractor of 19° (Δ15° condition) was classified as 'uncrossed'. A

trial was classified as 'crossed' when, for example, a 4° target was presented with a distractor of 169° (Δ −15° condition, and in this case crossing a cardinal axis).

Upper panels of Figure 5 show the probability that a cardinal axis was crossed for each Δ target-distractor condition. Note that for conditions with smaller Δ target-distractor (especially prevalent in Experiment 1), not a lot of trials occurred on which a cardinal axis was crossed. Middle panels show biases in the mean response when target and distractor orientations crossed or did not cross a cardinal axis. The original attraction (Figure 2, left panel) was more prevalent when only considering trials on which no cardinal axis was crossed (Figure 5, middle panels, dark teal lines). Interestingly, on 'crossed' trials, no attraction towards the distractor was observed, instead, the target was reported as further away from the distractor than its actual orientation. From Experiment 2 it could additionally be gleaned that once the target and distractor were distant enough, biases in the mean returned to baseline, and memory reports became veridical again. Lower panels plot the circular variance calculated separately for 'crossed' and 'uncrossed' trials. Variance was generally lower when the two orientations crossed a cardinal axis.

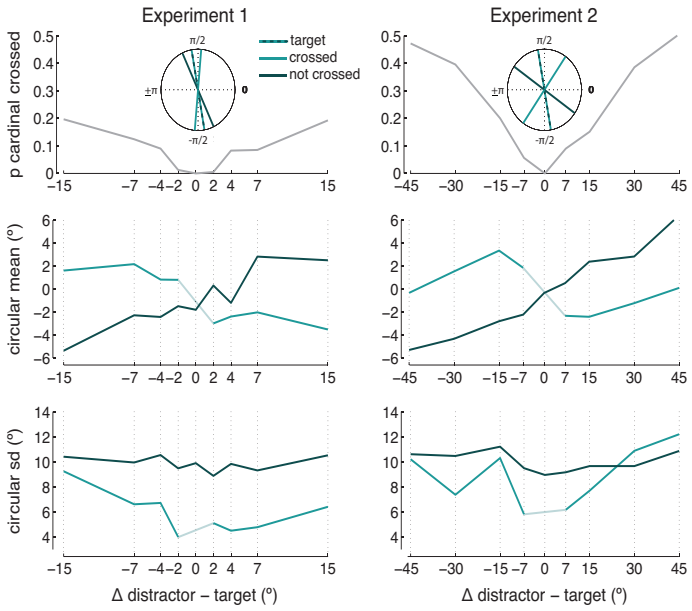


Figure 5. Cardinal crossing of target and distractor. The insets in the top panels show an example target orientation (in dashed light-and-dark teal) slightly counter-clockwise from vertical, and two possible scenarios in which the distractor either crossed (light teal) or did not cross (darker teal) a cardinal axis. Data were collapsed across participants, individual participant data can be found in Supplementary Figure 1.

Thus, cardinal axes impact memory for orientation, both in the absence and presence of a distractor. False memories due to an irrelevant distractor likely arise from a combination of both default biases (away from cardinal) as well as the influence of the distractor itself (attraction).

Condensing the orientation space. Before modeling the data, we condensed the 1–180° target orientation space onto a 0–90° space to make the data more robust. This is possible because the 1–180° target space (Figure 4) effectively consists of four repetitions of a cardinal-to-oblique (specifically: from 180° to 45°, from 180° to 135°, from 90° to 135°, and from 90° to 45°). We didn't assume any differences between default biases in clockwise or counterclockwise directions relative to cardinal, but were unsure about potential differences between horizontal and vertical cardinal axes. Hence, we collapsed the orientation space onto a condensed space running from vertical (0°) to horizontal (90°). Figure 6 shows the details of this procedure. After the data were collapsed, we again applied a sliding window analysis to obtain the circular standard deviation (Figure 7, top panels) and mean (Figure 7, bottom panels), this time for all the Δ target-distractor conditions (in color) in addition to the no-distractor baseline (in black). Circular parameters calculated across the original 1–180° space can be found in Supplementary Figure 2.

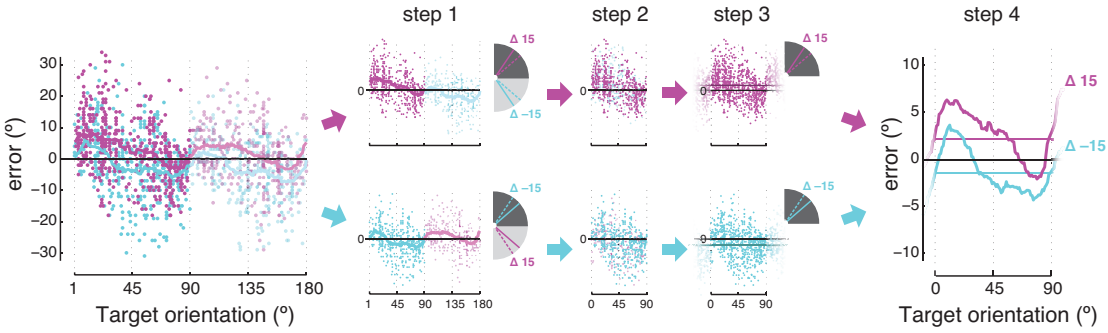


Figure 6. Mapping 1–180º orientation space onto a 1–90º space. On the far left the response errors (dots) and sliding window circular error (lines) are plotted for target-distractor conditions Δ –15 (in blue) and Δ 15º (in purple). Responses to targets < 90º are depicted in full color, whereas responses to targets > 90º are depicted transparently. The first step to consider before flipping the full 1–180º space onto a 1–90º space, is that (following the purple arrows) a Δ 15º condition for targets < 90º corresponds to a Δ –15º condition for targets > 90º. Vice versa (following the blue arrows), Δ –15º condition for targets < 90º corresponds to a Δ 15º condition for targets > 90º. This mirroring of conditions is represented by the half-circles that represent an orientation space from 1–180º (dark grey represents targets < 90º, light grey represents targets > 90º). The data depicted in the scatter plots highlights the data to be combined in the next step. The second step overlays data from targets < 90º with data from targets > 90º from the Δ condition with opposing sign. Data from targets > 90º are re-aligned by calculating 180 minus each target (º) and flipping the sign of each error (º). The third step is to expand the orientation space a little on both sides of cardinal (aligned to the mean response in each condition) in order to accurately calculate the sliding window. The fourth and final step shows the mean circular error (here denoted 'error' for short) calculated across the 20º sliding window. For illustrative purposes we chose to demonstrate this procedure by taking the Δ–15º and Δ15º target-distractor conditions as an example, but this procedure was applied in the same manner for all other conditions.

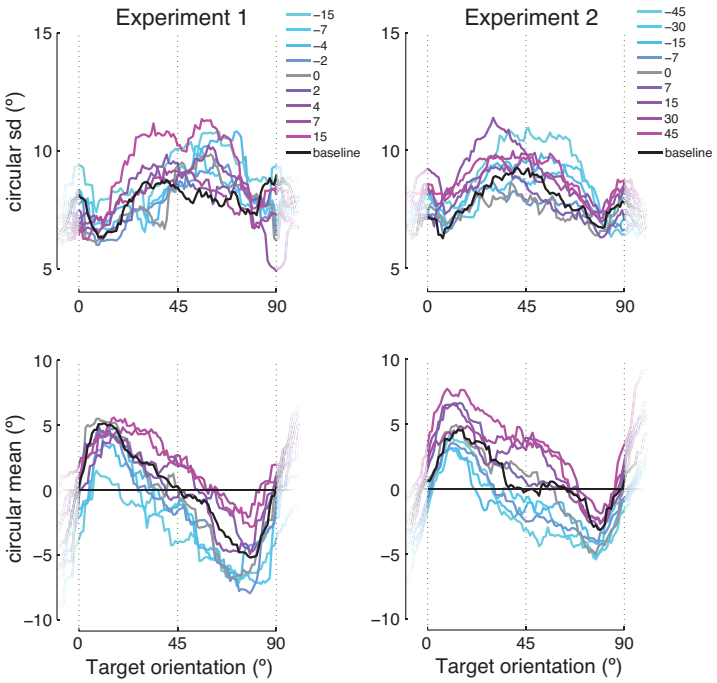


Figure 7. Circular parameters. SD (top panels) and mean (bottom panels) calculated within a 20° sliding window, plotted against target orientation condensed onto a 0-90° space. Variance is generally larger for larger Δ target-distractor conditions (i.e. brighter blue and purple lines). Mean responses show a graded shift in the direction of the distractor. For example, when a distractor is rotated counter clockwise relative to a target (blue lines) the responses are shifted in counterclockwise direction.

Model 1: weighted averaging. The principle behind the first model is simple: we assumed a default bias for both the target and the distractor (based on the no-distractor baseline), and use a linear combination rule to estimate the response. Formally, the first model is defined as:

$$p_{combined} = pT \times (1\text{-}w) \; + pD \times w \qquad (1)$$

Where $P_{combined}$ is the reported percept, which is a combination of the perceived target (pT) and the perceived distractor (pD), each with a respective weighting (w), the weights summing to 1. This idea is schematically depicted in Figure 8. The left panel shows a simulated baseline (in black) over the full 1–180° orientation space, with repulsion away from the cardinal axes. While we cannot know the representation of the distractor (it is never probed) the first model assumes it is described by the same function, including a bias away from cardinal. In our schematized depiction, the distractor is rotated by

10º relative to the target, and plotted in red as a function of the target orientation (i.e. it is a shifted replica of the black curve). In blue is a linear combination of the two, with a weight of 75% for the target, and a weight of 25% for the distractor. The right panel shows the simulated $P_{combined}$ for distractors ranging from Δ –45º to 45º (as in Experiment 2).

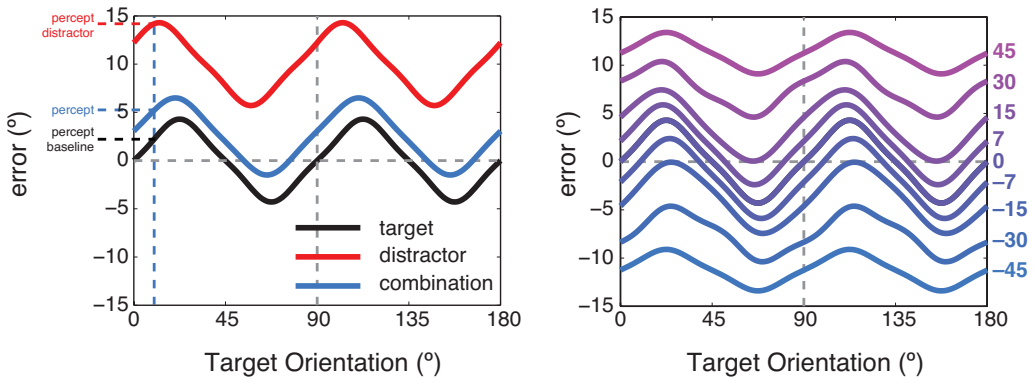


Figure 8. Schematic of weighted averaging model. The left panel shows default biases for a simulated baseline (black) and a simulated distractor rotated 10º clockwise relative to the target (red). The red curve is shifted to the left relative to the black curve, indicating the clockwise rotation of the distractor relative to the target. The blue dashed line indicates an example trial during which a 10º target was presented. On this trial, the distractor was presented at 10º+10º = 20º. The percept of both target and distractor (given default bias) can be read out from the y-axis. The target was perceived as 10º + a ~2º bias away from cardinal (i.e. ~12º) and the distractor was perceived as 20º + a ~4º bias away from cardinal (i.e. ~24º). Given a simple linear combination rule assigning a weight of 75% to the target and 25% to the distractor the response on this trial was perceived as 0.75 * 12 + 0.25 * 24 = 15º. The right panel shows simulated predictions from the weighted averaging model over the Δ target-distractor conditions from –45 through 45º (as in Experiment 2).

The model was fit to the circular mean of the response, condensed onto a 1–90º orientation space (as shown in Figure 7, lower two panels), and this data is re-plotted as empty circles in Figures 9 and 10 (for Experiments 1 and 2 respectively). Additionally, these two figures plot pT in black (note that this is identical to no-distractor baseline), and the predicted $P_{combined}$ in colored lines. In Experiment 1 the predicted response resulted from a target weighted by 81.2%, and a distractor weighted by 18.8%. In Experiment 2 the influence of the distractor was considerably less according to the model, namely 5.9%,

while the target was weighted with 94.1%. The fits were slightly better for Experiment 1 with an $R^2 = 0.89$ ($\pm$ 0.017 SE) than for Experiment 2 ($R^2 = 0.79 \pm 0.03$ SE).

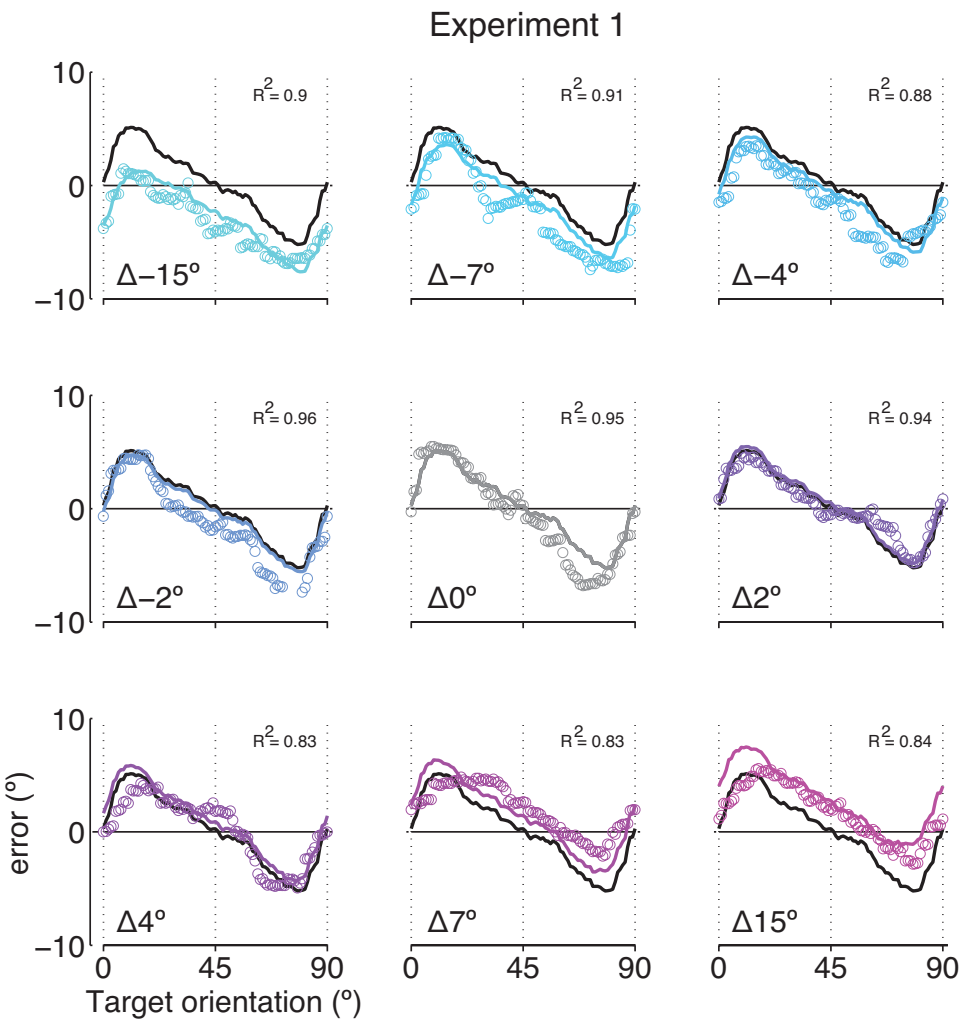


Figure 9. Weighted averaging model fits for Experiment 1. Real (circular) mean response errors are plotted against the target orientation in empty colored circles. Overlaid solid colored lines represent the fitted errors for the various Δ target-distractor conditions according to the weighted averaging model. Baseline data in the absence of a distractor (in black solid lines) were used to model target and distractor percepts.

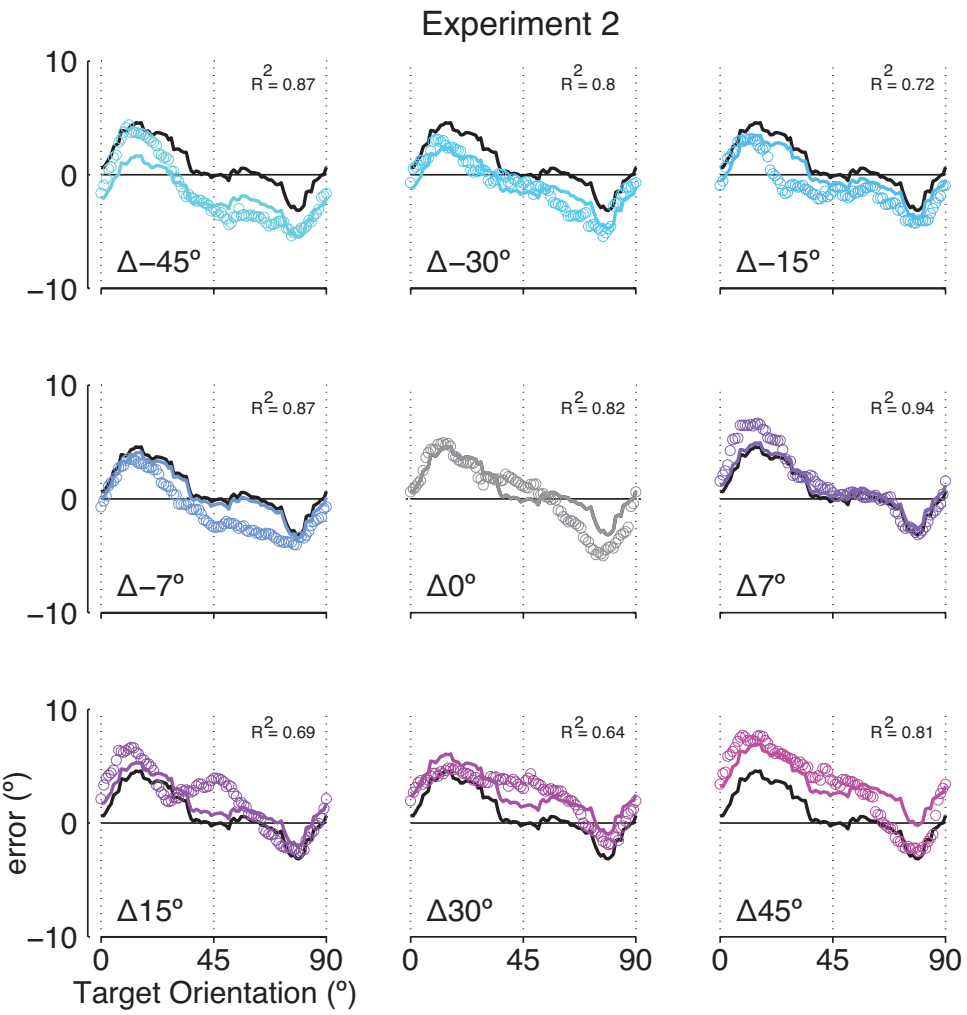


Figure 10. Weighted averaging model fits for Experiment 2. Observed circular mean errors are plotted against the target orientation (empty colored circles) with overlaid the fitted errors for the various Δ target-distractor conditions according to the weighted averaging model (solid colored lines). Baseline data in the absence of a distractor (black solid lines) were used to model target (*pT*) and distractor (*pD*) percepts, and here they are shown to represent *pT*.

Error residuals were calculated by subtracting the model predictions from the real data in each Δ target-distractor condition (Figure 11). Residuals appeared to show some systematic deviations, with a shape reminiscent of the variance. Also (and especially for Experiment 2), the model underestimated the shift in responses for larger Δ target-

distractor. A possible way to improve the model is to take the response variance into account. After all, while we have assumed equal variances for target and distractor percepts this is unlikely the case. For one, variance depends on the target orientation, with more precision around the cardinal axes. Given that the target and distractor often differed from one another in their orientation, their variances should therefore also differ.

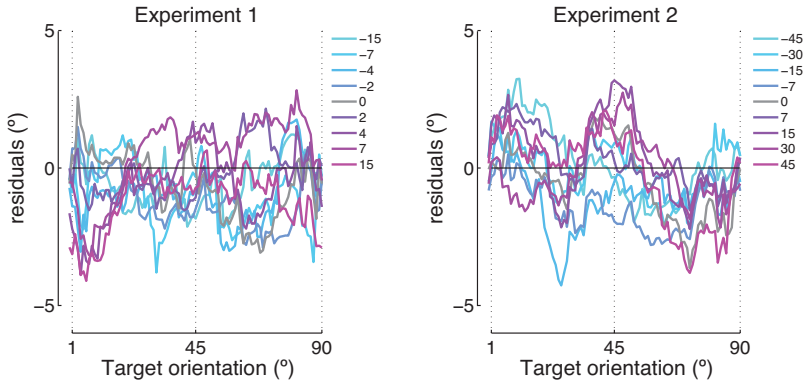


Figure 11. Residuals from the weighted averaging model. Residuals were calculated by subtracting the model estimates from the real data.

To obtain a target weights contingent on the target reliability we calculated:

$$w^T = \frac{(1-\alpha)\, r_T}{(1-\alpha)\, r_T + \alpha r_D} \tag{2}$$

Where w^T is the target weight (the weight of the distractor $w^D = 1 - w^T$), α represents the model parameter (variance contingent weight), and r_T and r_D representing the target and distractor reliability respectively. The target reliability is calculated as follows:

$$r_T = \frac{1}{\delta 2(\theta T)} \tag{3}$$

With $\sigma^2\ (\theta^T)$ representing the variance of the target orientation. The target reliability was calculated based on the circular variance during no-distractor baseline trials (Figure 7, top two panels, solid black curve) for each possible target orientation. The distractor reliability (r_D) was a shifted replica of the target reliability. Fits for all conditions are plotted together

in Figure 12, allowing direct visual comparison with weighted averaging fits that were not variance contingent (shown in the inserts). Individually plotted Δ target-distractor condition fits can be found in Supplementary Figures 3 and 4 (for Experiments 1 and 2 respectively).

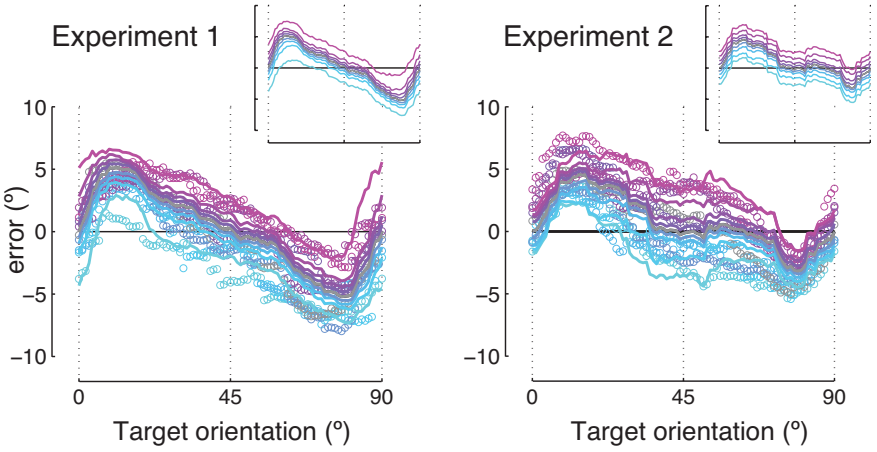


Figure 12. Weighted averaging model fits. Estimated mean responses from the variance contingent weighted averaging model are plotted in solid colored lines, with empty colored circles representing real response error means. Fits from the weighted averaging model (Figures 9 and 10) are replotted in the inserts for comparison. Colors correspond to conditions as depicted in all other figures.

Target weights using the variance contingent model were 82.3% and 94.6% for Experiments 1 and 2 respectively. The quality of the fits with the variance taken into account (Experiment 1: $R^2 = 0.88 \pm 0.02$ SE, Experiment 2: $R^2 = 0.79 \pm 0.03$ SE) hardly differed from those when the variance was not taken into account. Where they did differ, they were slightly worse.

In sum, a simple linear averaging model did fairly well describing responses to a target in the presence of a distractor, with little to no advantage when taking the variance into account. The weighting of a target relative to a distractor depended on the range of Δ target-distractor conditions used, with distractors carrying less weight in Experiment 2 which used a larger range.

Model 2: Decision and integration. While the first model does a reasonable job explaining target responses in the presence of a distractor, it is agnostic about the existence of the default bias. Moreover, the model assumes that, on average, the target and distractor are coded with the same amount of variance, while one could reasonably expect an irrelevant distractor to be represented with less precision than the target. To take these factors into account, we proposed a second model that relies on a decision stage as well as an integration stage during which two distributions of unequal variances are combined.

First, a target distribution was defined as follows:

$$d_T = \frac{1}{\sqrt{2\pi\delta_T^2}} \exp\left(-\frac{(x-\mu_T)^2}{2\delta_T^2}\right) \quad (4)$$

Where the variance of the target (σ^2_T) was taken to be the mean variance during no-distractor baseline trials, which was 8.47° and 8.17° for Experiments 1 and 2 respectively (collapsed across all participants). Distributions were defined for every possible target orientation (μ_T).

Next, the model assumed a decision concerning the orientation of a target relative to the cardinal axes. This is schematically depicted in the upper left panel of Figure 13, where a target distribution (d_T in grey) around example target orientation (T°) proximal to a cardinal axis (0°) was split along this cardinal axis. Depending on whether participants chose the correct (in green) or incorrect (in red) side of the distribution on a given trail, the target distribution was cropped according to $d^T(-\infty,0) = 0$ or $d^T(0,\infty) = 0$ respectively. Of course, the probability of choosing the correct side is always larger as long as T° ~= 0. Summing the cropped distributions for correct and incorrect choices weighted by their respective probabilities (area under the curve) across all possible target orientations, a mean response away from cardinal emerged (Figure 13, top right panel).

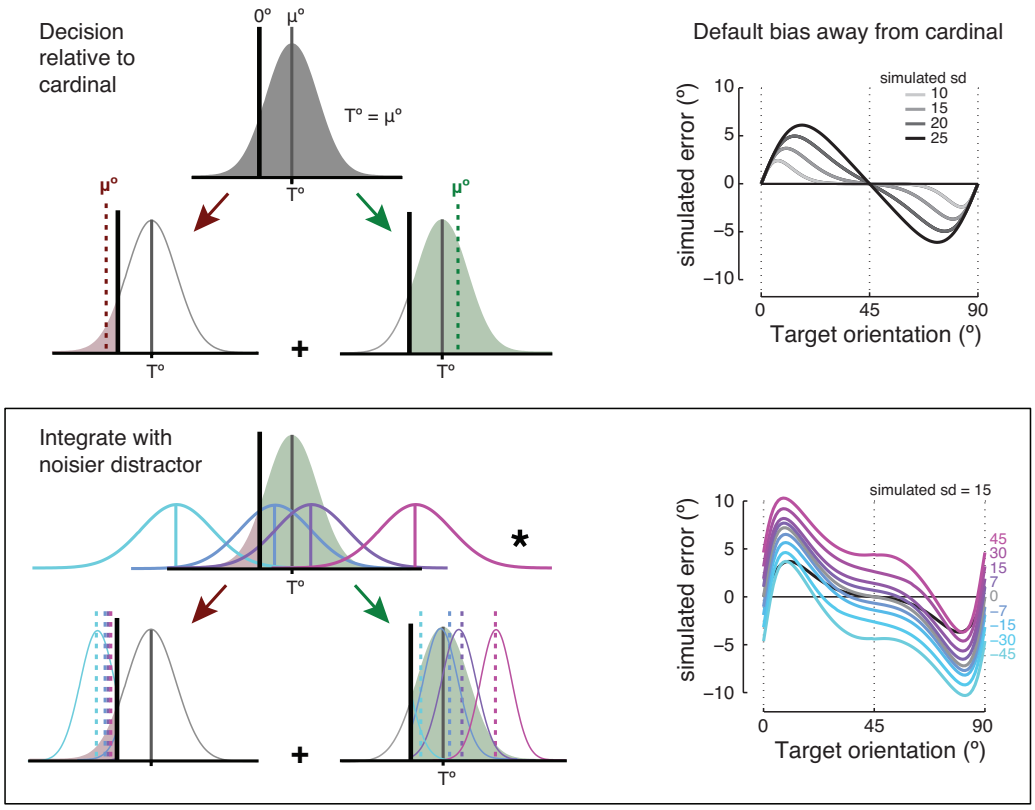


Figure 13. Schematic of the decision and integration model. Upper panel: The internal representation of example target orientation T° has a certain amount of noise (upper left, in grey). When T° is proximal to a cardinal axis, a decision is made about the identity of T° relative to that cardinal axis, with all information from the not-chosen side being dismissed. Most of the time this decision will correctly reflect the true side of the cardinal (in green), but on some proportion of trials T° will be estimated to lie at the incorrect side of cardinal (red). The proportional sum of correct and incorrect choices for every possible target orientation gives mean response estimates showing a default bias away from cardinal (upper left panel – these are simulated data, displaying the outcomes for target distributions with varying standard deviations). Lower panel: When a distractor is presented whose variance is larger than that of the target, the post-choice cropped distribution undergoes point-by-point multiplication with the distractor (upper left). Here, we showed four possible distractor distributions in blue-to-purple (clockwise-to-counter clockwise relative to the target) for illustrative purposes. The resulting distributions for correct and incorrect choices are shown on the lower left, again depicted in blue-to-purple for the four possible example distractors (resulting distribution means are depicted in blue-to-purple dashed lines). Summing the correct and incorrect distributions weighted by their respective probabilities, for every possible target orientation resulted in the (simulated) means on the lower right, plotted against all possible target orientations to reflect the model's predictions.

After cropping during the decision stage, the target distribution undergoes point-by-point multiplication with a distractor distribution:

$$d_D = \frac{1}{\sqrt{2\Pi\delta_D^2}} \exp\left(-\frac{(x-\mu_D)^2}{2\delta_D^2}\right) \quad (5)$$

The target and distractor distributions were assumed to have unequal variances, namely $\delta_D = \delta_T \times w$, with w being a free parameter determining the scaling of the distractor standard deviation relative to the standard deviation of the target. Integration of the cropped target and wider distractor distributions is schematically depicted in the bottom left panel of Figure 13. Summing the multiplied distributions by their respective proportions yielded the predictions shown in the lower right panel of Figure 13. One interesting prediction of this model was a stronger repulsion from cardinal when both target and distractor orientations were the same ($\Delta 0^o$ condition), compered to the repulsion from cardinal during baseline trials.

Data with fits from the decision and integration model overlaid are shown in Figure 14 (conditions are shown individually in Supplementary Figures 5 and 6 for Experiments 1 and 2 respectively). In Experiment 1 the model's scale parameter estimated that the distractor standard deviation was 1.88 larger than that of the target. In Experiment 2 the distractor standard deviation was estimated to be 3.8 times that of the target, indicating that (similar to the weighted averaging model) the influence of the distractor was considerably less in the second experiment. On average the explained variances were R^2 = 0.87 ($\pm$ 0.02 SE) and R^2 = 0.8 ($\pm$ 0.02 SE) for Experiments 1 and 2 respectively. Thus, while the decision and integration model fits the data from Experiment 2 slightly better than the weighted averaging model, for Experiment 1 the weighted averaging model performs better.

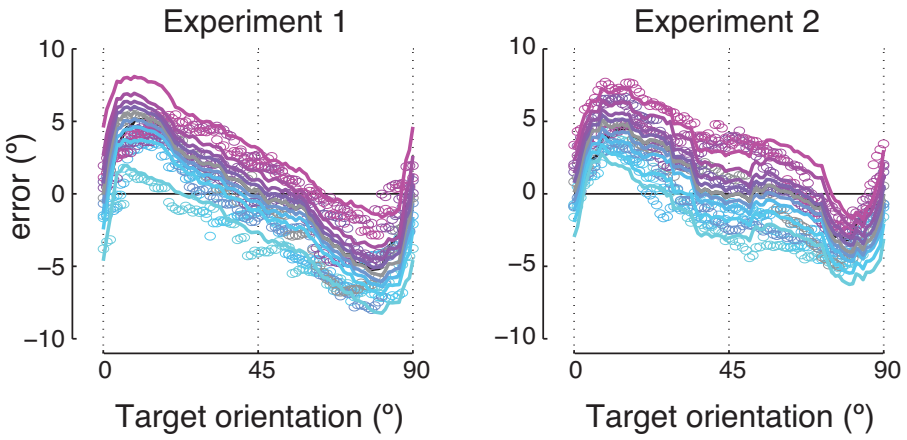


Figure 14. Decision and integration model. Fits from the decision and integration model are plotted in solid colored lines, with empty colored circles representing the observed mean response errors, calculated within a sliding window of 20°. Colors correspond to conditions as depicted in all other figures.

Discussion

When observers replicated an orientation from memory, responses were biased towards an irrelevant distractor orientation shown during the delay (Rademaker et al., 2015). Such attraction towards irrelevant information has been observed during the maintenance of other simple visual features as well (Dubé, et al., 2014; Huang & Sekuler, 2010; Nemes, et al., 2011; 2012). Here we attempted to uncover possible computations that might bring about this kind of false memory for orientation.

First, it should be noted that exploring memory for orientation revealed not one, but two sources of systematic response errors: The first being an attraction towards the distractor (Rademaker et al., 2015), the second being the non-veridical reports observed in the absence of a distractor, consisting of a repulsion away from cardinal (Figure 4, lower panels). We dubbed the second phenomenon a 'default bias'. This default bias in the mean response might constitute a 'false percept', occurring irrespective of the presence of a memory component. However, it is yet unclear to what extent this bias might actually reflect a systematic misremembering of orientation information. Thus, while we use the

term 'false memory' to refer to attraction towards a distractor, we cannot exclude that the 'default bias' might also be a systematic distortion of memory.

The first model quantified memory attraction by assuming a simple linear weighted average of the target and distractor orientations. This weighted averaging model predicted the observed responses well, providing a clear and simple solution for false orientation memory in the face of a distractor. Moreover, this model can generalize easily to false memories for other visual features, such as memory for spatial frequency (Dubé et al., 2014). However, a weakness of the model is that it simply assumes the default bias in the mean response without explaining it, and is thus completely agnostic about its origin. Moreover, the model assumes that, while distractor representations are shifted versions of target representations, they are equal in all other respects. This might not be true, as the distractor was neither attended nor remembered, and conceivably represented with a larger overall variance than the target. While the variance contingent weighted averaging model accounted for differences in variance due to the oblique effect, it still assumed target and distractor variances to be equal otherwise. This could be part of the reason why the variance contingent model did not yield an improved fit compared to its more basic, non-variance contingent counterpart. Additionally, the estimated baseline variance that fed into the variance contingent model was noisy: Circular variance is less robust when calculated across sparse samples than the circular mean, and with 100 baseline trials per participant less than 9 data-points were collected per target orientation.

The second model quantified memory attraction as arising from integration of a cropped target distribution and a noisier distractor distribution. The role of the cardinal axes in orientation memory inspired the conception of this decision and integration model, and the decision stage in particular. Firstly, a category decision is made regarding where the target orientation is relative to cardinal, allowing the target distribution to be cropped by discounting all information that does not align with this choice. The decision might occur at the perceptual stage or during memory, and can by itself predict the default bias observed in the mean responses (Figure 13, top panels). Secondly, the model assumes a

distractor distribution that is noisier than the target distribution, and multiplicative integration of the two. The resulting memory attraction predicted by the model fit the observed behavior about equally well as the predictions made by the weighted averaging model, but with the added advantage of being able to account for the default bias in the response mean. Note that the model has the additional potential to explain the oblique effect, as cropped target distributions close to cardinal should reduce the variance.

The decision and integration model lends itself well to modeling memory for orientation where the cardinal axes impose natural categorical boundaries that can be utilized to make categorical decisions. In a similar vain, a categorical model has been suggested for location memory (Huttenlocher, Hedges, & Duncan, 1991), which has similarly clear-cut spatial reference frames. However, it is less obvious how the decision and integration model might work for other features (like color or spatial frequency) where the possible decisional parameter space is less straightforward.

It's been previously shown that participants remembering and replicating a direction of motion exhibited no consistent response biases (Blake et al., 1997), despite having a cardinal reference frame similar to orientation. Nevertheless, this study uncovered strong idiosyncrasies, with some participants showing biases away from cardinal, while others showed biases towards cardinal. This raises an interesting point, as decisional strategies might be highly personal, and could result in qualitatively different biases from one person to the next. Indeed, cognitive and decisional processes are known to influence biased reports of basic visual features in memory (Blake et al., 1997; Park, Rademaker, & Tong, 2014; Rademaker et al., 2015), and neither of the two models discussed here have taken such idiosyncrasies into account. A logical next step might therefore be to collect large sets of data from individual participants to test our models against. This might also further our understanding of the repulsion found by Rademaker at al. (2015), something that cannot exist when assuming linear weighted averaging as an explanation for false orientation memory. While the decision and integration model in its current form can

also not predict repulsion biases, one could easily envision some type of flexible updating of the decision strategy to expand this model and its declarative power.

Both models indicated that the distractor influence differs between Experiments 1 and 2, its influence reduced in Experiment 2. This is indicated by the smaller estimated weight for the weighted averaging model, and a larger SD scaling factor for the decision and integration model. Clearly, the larger range of target-distractor differences tested in Experiment 2 reduces the influence of the distractor, resulting in the apparent range invariance of the attraction effect. However, neither model actually explains this range invariance in a quantifiable way. A previously suggested possibility is that the range of possible distractors is learned and used as a prior to inform behavioral outcomes (Rademaker et al., 2015). Although our time-resolved analyses showed attraction from very early on (after a minimum of nine trials), they might nevertheless lack sensitivity. It has been shown that stimulus contingencies can be learned very fast, within very few trials (Lages & Treisman, 1998; Lages & Paul, 2006). Additionally, our participants performed several practice trials, which may have allowed them to discover the task structure (and building of prior expectations) well before even starting the experimental runs.

One more finding we would like to account for, but that is currently missing from our models, is that attraction has been found to return to baseline once the target and distractor became sufficiently distant (Nemes et al., 2011; 2012; Van der Stigchel et al., 2007). While such return to baseline was not apparent in the data presented here, it is obvious that a return to baseline is mandatory, as once a distractor is orthogonal to a target no more attraction is possible. Thus, future models should be able to account for the circular wrapping of variables such as orientation. Interactions between targets and distractors depend on their distance, as previously suggested by the idea of channel interactions (Magnussen & Greenlee, 1999; Magnussen, 2000), as well as more recently in continuous attracter state models (Sreenivasan, Curtis, & D'Esposito, 2014). In the models presented here such interactions might be implemented by making the weights (weighted

averaging model) or scaling factor (decision and integration model) contingent on the target-distractor difference.

What might have brought about the clear differences between crossed and uncrossed trials demonstrated in Figure 5? While these observations implicated that the importance of cardinal axes extended beyond default biases, including also the way in which a distractor impacts a target, is this really the case? When looking at the data (Figure 7, lower panels, as well as the unfilled circles in Figures 9, 10, 12, and 14) it is clear that under some circumstances the default bias dominates the final response, irrespective of the target-distractor difference. For example, although a distractor rotated 45° clockwise relative to a target will pull that target towards its clockwise direction, if the target orientation was between approximately 70° and 85° the response will nevertheless be biased towards counter clockwise.

Thus, a more parsimonious explanation for the observed differences between 'crossed' and 'uncrossed' trials (Figure 5) comes from considering the sampling of those trials in the presence of a default bias. Note that, as long as a target and distractor differed by less than 45°, axis crossing is more likely for target orientations closer to cardinal (Figure 5, top panels). And this likelihood increases as target-distractor differences become smaller. The lack of an attraction effect on 'crossed' trials is likely an artifact of the target orientations for which it holds true that a cardinal axis was crossed. This sampling issue is graphically demonstrated in Figure 15, showing for two example conditions (Δ –7° and Δ 45°, replotted from the lower right panel of Figure 7) which trials were included when calculating parameters for 'crossed' trial instances (indicated by the shaded regions). For distractors rotated 7° counterclockwise relative to a target (Δ –7°), a cardinal axis is only crossed when the target orientation was between 0° and 6° (blue shaded region). Responses to these target orientations happen to be strongly biased away from cardinal, making it appear as if the counterclockwise distractor pushed the responses in a clockwise direction (i.e. a repulsion) when in effect that is not the case. In comparison, when the distractor was rotated 45° clockwise relative to the target, a cardinal axis was crossed for

target orientations of 45º through 90º (shaded purple region) –eliminating the dominant role of the default bias.

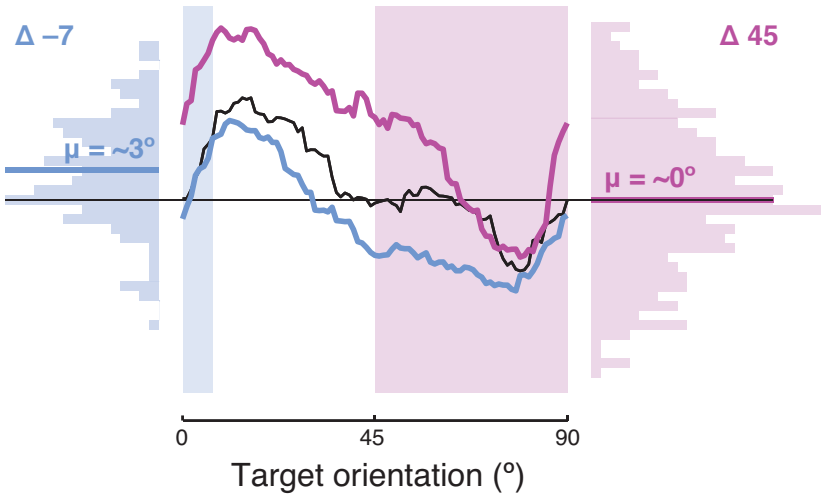


Figure 15. Selective sampling of target orientations for 'crossed' and 'uncrossed' trials The default bias away from cardinal was largely responsible for the apparent repulsion during 'crossed' trials in the Δ –7º target-distractor condition (in blue). Shaded blue regions represent the only target orientations in this condition for which a distractor was presented on the 'crossed' side of a cardinal axis, as well as the error histogram from the corresponding 'crossed' trials. For all other target orientations in this condition the distractor did not cross a cardinal axis. When the Δ target-distractor is larger (Δ 45º in this example) biases in the mean response are calculated over a much larger window, resulting in an error histogram that is no longer dominated by the repulsion from cardinal. Note also how this Δ 45º condition demonstrates why only considering 'uncrossed' trials would lead to a larger attraction then when considering all trials: the default repulsion from cardinal is now in a direction favoring the distractor.

Similarly, the reduction in response variance for crossed compared to uncrossed trials (Figure 5, lower panels) could simply reflect the oblique effect: Axis crossing is more likely for orientations closer to cardinal, which by virtue of their proximity to cardinal are already less noisy to begin with. This notion is supported by smaller differences between 'crossed' and 'uncrossed' standard deviations for larger Δ target-distractor conditions. In sum, the apparent influence of cardinal axes on memory attraction is probably largely

spillover from default biases (i.e. the influence of cardinal axes on responses in the absence of a distractor).

Finally, false memories from task-irrelevant information can arise for many other possible reasons. For example, memories can be biased towards the prototypical or average value of stimuli seen in the recent past (Wilken & Ma, 2004; Huang & Sekuler 2010). It's been suggested that such reliance on task-irrelevant information could be adaptive, serving as a supplement to imperfections in memory. The idea is that with degrees of uncertainty people might rely more on these types of information (Freyd & Johnson, 1987; Spencer & Hund, 2002). Thus, future attempts to model false memories might want to account for these other components in addition to the influence exerted by irrelevant distractors.

References

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin, 78*(4), 266–278.

Bennett, P.J. & Cortese, F (1996). Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. *Vision Research, 36*(2), 233–238.

Blake, R., Cepeda, N.J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 353–369.

Dubé, C., Zhou, F., Kahana, M.J., & Sekuler, R. (2014). Similarity-based distortion of visual short-term memory is due to perceptual averaging. *Vision Research*, *96*, 8–16.

Freyd, J. J., & Johnson, J. Q. (1987). Probing the time course of representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(2), 259–268.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: two classes of attractors at work. *Journal of Vision*, *10*(2):24, 1–27.

Huttenlocher, J., Hedges, L. V., & Duncan, S. (1991). Categories and particulars: prototype effects in estimating spatial location. *Psychological Review*, *98*(3), 352–376.

Lages, M., & Treisman, M. (1998). Spatial frequency discrimination: visual long-term memory or criterion setting? *Vision Research, 38*(4), 557–572.

Lages, M., & Paul, A. (2006). Visual long-term memory for spatial frequency? *Psychonomic Bulletin & Review*, *13*(3), 486–492.

Magnussen, S., Greenlee, M.W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, *31*(7-8), 1213–1219.

Magnussen, S., & Greenlee, M.W. (1992). Retention and disruption of motion information in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 151–156.

Morgan, M.J., Watamaniuk, S.N., & McKee, S.P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research*, *40*(17), 2341–2349.

Nemes, V. A., Parry, N. R., Whitaker, D., & McKeefry, D. J. (2012). The retention and disruption of color information in human short-term visual memory. *Journal of Vision*, *12*(1):26, 1–14.

Nemes, V. A., Whitaker, D., Heron, J., & McKeefry, D. J. (2011). Multiple spatial frequency channels in human visual perceptual memory. *Vision Research*, *51*(23), 2331–2339.

Park, Y.E., Rademaker, R.L, & Tong, F. (2014). Both variations in Perceptual sensitivity and decisional response bias contribute to visual working memory performance. Journal of Vision 14(10), 1375–1375.

Phillips, W.A. (1974) On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*(2), 283–290.

Rademaker, R.L., Bloem, I.M., De Weerd, P., & Sack, A.T. (2015). The impact of interference on short-term memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance*.

Spencer, J.P., & Hund, A.M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology General*, *131*(1), 16–37.

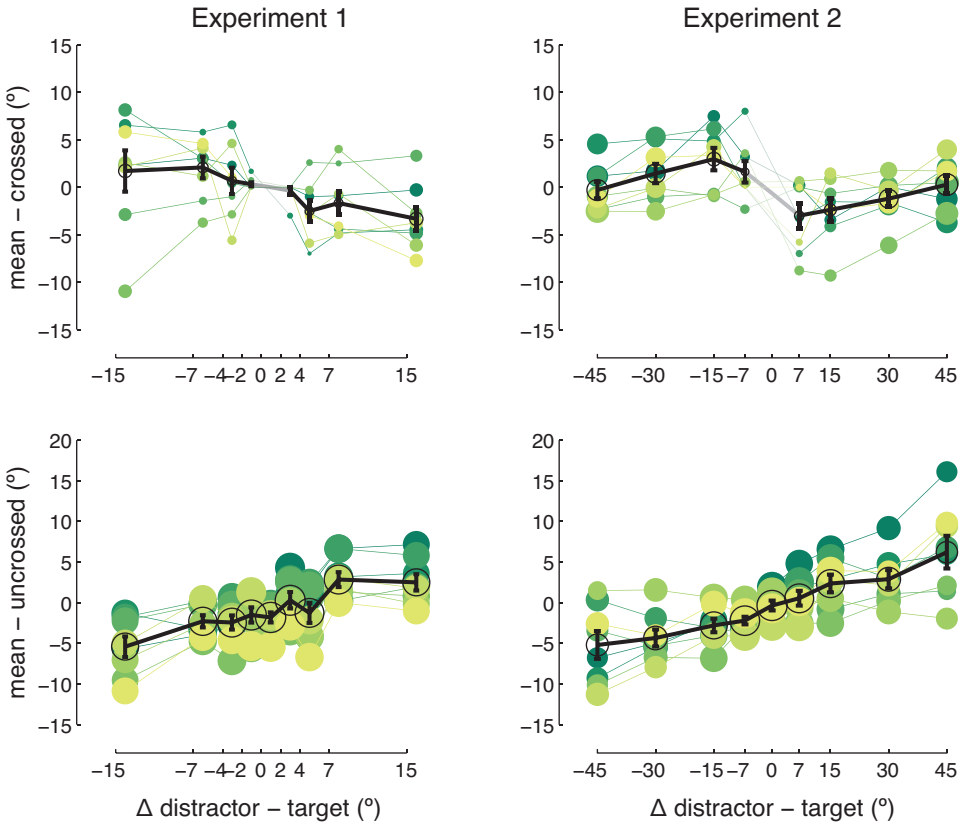
Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89.

Van der Stigchel, S., Merten, H., Meeter, M., & Theeuwes, J. (2007). The effects of a task-irrelevant visual event on spatial working memory. *Psychonomic Bulletin & Review*, *14*(6), 1066–1071.

White, J.M., Sparks, D.L., & Stanford, T.R. (1994). Saccades to remembered target locations: An analysis of systematic and variable errors. *Vision Research, 34*(1), 79–92.

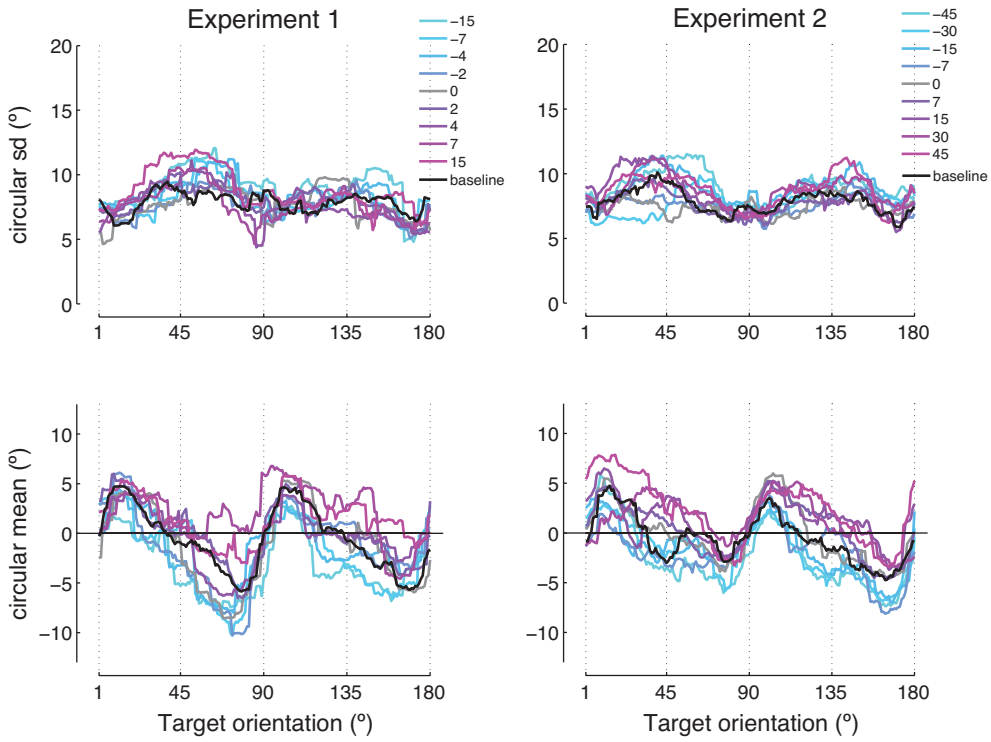
Wilken, P., & Ma, W. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12), 1120–1135.

Supplementary Figure 1



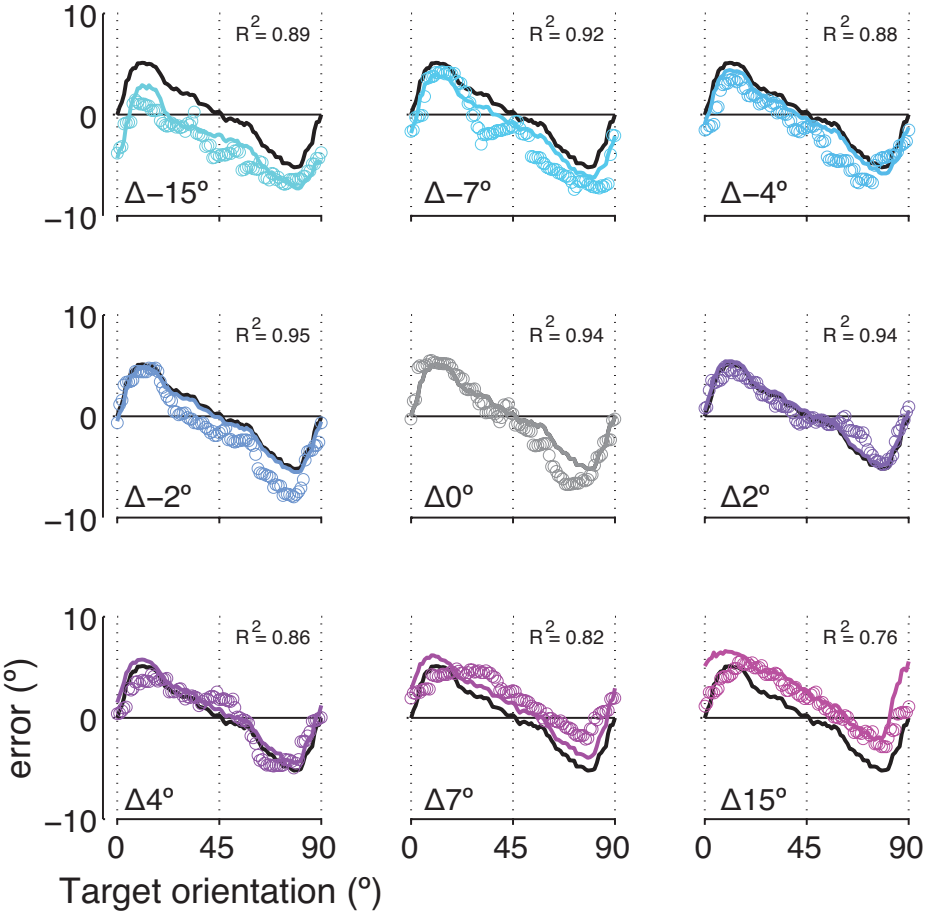
Supplementary Figure 1. Biases in the response mean when a target and distractor did, or did not cross a cardinal axis. Upper panels show that memories were biased away from the distractor, if that distractor was presented on the other side of a cardinal axis. Lower panels show that responses to a target orientation were attracted towards a distractor orientation when no cardinal axis was crossed between the two. Individual participants are represented in shades of green, with dot size representing the number of trials underlying each estimate (larger dots representing more trials / robust estimates).

Supplementary Figure 2



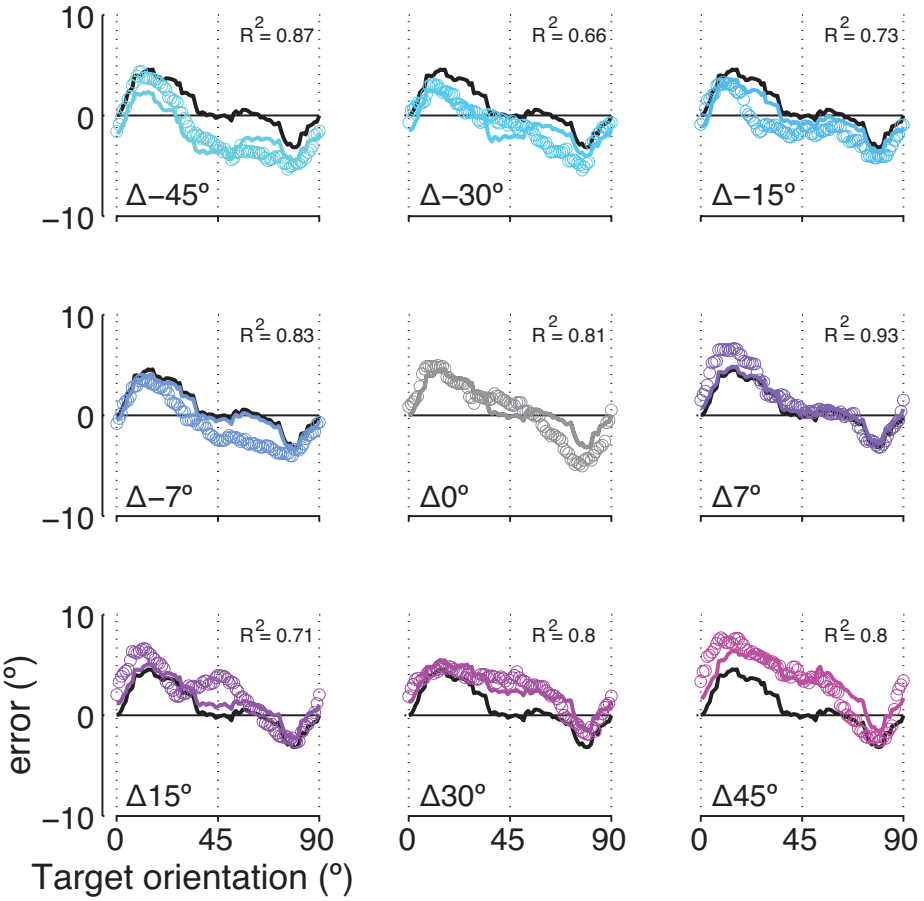
Supplementary Figure 2. Circular parameters from all experimental conditions (1–180°). Top panels show the circular SD (in °) measured over a 20° sliding window for all experimental conditions (see legend). Variance is generally larger for larger Δ target-distractor conditions (i.e. brighter blue and purple lines). Bottom panels show the circular mean similarly calculated over a sliding window – all are plotted against the target orientation. Bottom panels show a graded shift of the mean response in the direction of the distractor. For example, when a distractor is rotated counter clockwise relative to a target (blue lines) the responses are shifted in counterclockwise direction. Vice versa, a clockwise shift is observed for responses under conditions during which the distractor was clockwise relative to the target (purple lines).

Supplementary Figure 3



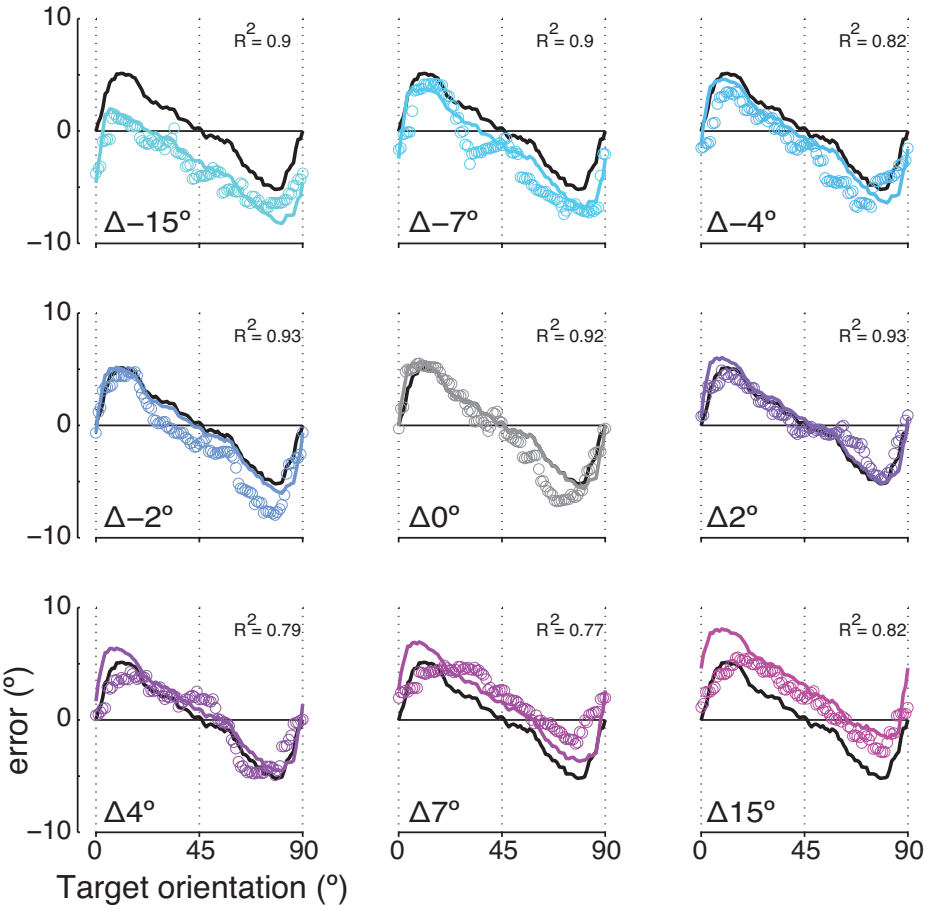
Supplementary Figure 3. Variance contingent weighted averaging model fits for Experiment 1. Real response errors are plotted against the target orientation in empty colored circles. Overlaid solid colored lines represent the fitted errors for the various Δ target-distractor conditions according to the weighted averaging model. Baseline data in the absence of a distractor (in black solid lines) were used to model target and distractor percepts.

Supplementary Figure 4



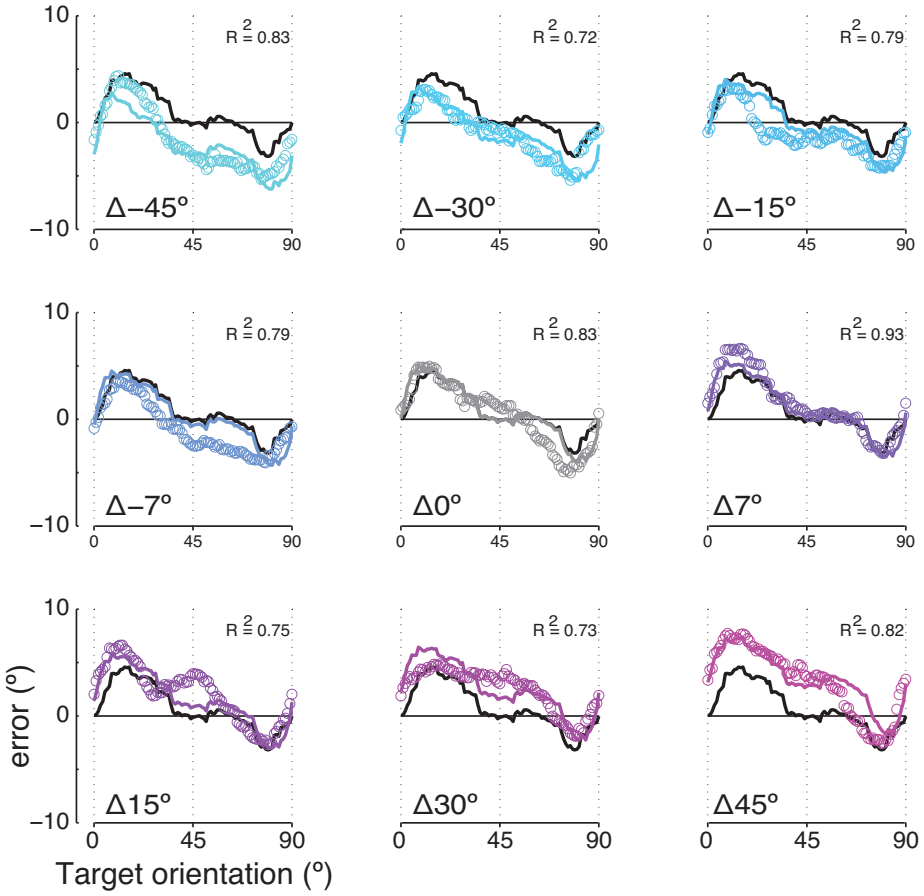
Supplementary Figure 4. Variance contingent weighted averaging model fits for Experiment 2. Real response errors are plotted against the target orientation in empty colored circles. Overlaid solid colored lines represent the fitted errors for the various Δ target-distractor conditions according to the weighted averaging model. Baseline data in the absence of a distractor (in black solid lines) were used to model target and distractor percepts.

Supplementary Figure 5



Supplementary Figure 5. Decision and integration averaging model fits for Experiment 1. Real response errors are plotted against the target orientation in empty colored circles. Overlaid solid colored lines represent the fitted errors for the various Δ target-distractor conditions according to the decision and integration model.

Supplementary Figure 6



Supplementary Figure 6. Decision and integration averaging model fits for Experiment 2. Real response errors are plotted against the target orientation in empty colored circles. Overlaid solid colored lines represent the fitted errors for the various Δ target-distractor conditions according to the decision and integration model.

Chapter 4

Decay of visual short-term memory as a function of time

Manuscript in preparation:

Rademaker R.L., Park, Y., Sack A.T., Tong, F.

Decay of visual short-term memory as a function of time.

Abstract

People can maintain visually precise information in working memory over a range of many seconds, in a manner that appears robust to the duration of delay. Once visual information is encoded into working memory, one might expect that it should inevitably begin to degrade over time, as this actively maintained information is no longer tethered to the original perceptual input. However, both previous and recent studies have suggested that items can be stably maintained in working memory with negligible loss of precision over time, especially for attentionally prioritized items. Here, we rigorously examined this issue by evaluating working memory for single central presentations of an oriented grating, color patch, or face stimulus, across a range of delay intervals (1, 3, 6 or 12s). We applied a mixture-model analysis to distinguish changes in memory precision from changes in the frequency of outlier responses that resemble random guessing. For all three types of stimuli, we observed a modest but highly reliable decline in the precision of working memory as a function of temporal delay, as well as an increase in guess-related responses for colored patches. Our results demonstrate that visual working memory is far from lossless. Although basic visual features and complex objects can be maintained in a quite stable manner over time, these working memory representations are still subject to stochastic noise and gradual decay.

Introduction

Visual information enters the eyes in a constant stream of brief fixations. Often, this information is retained over variable delays. Depending on the task at hand such memories could be, for example, integrated with other visual evidence to construct a more complete visual percept of the surrounding world, or used to compare one visual item to another. Irrespective of the task, visual information must be stored with high fidelity, and protected against decay over time. Here we utilize a recall paradigm to address a question of perennial interest in the area of human memory – the effect of time on visual short-term memory representations.

The question whether visual memories degrade over time, while seemingly simple, has been grounds for debate over a good many years. One option is that memories deteriorate gradually as a function of time. This could happen in one of two ways, the first is stochastic and assumes a random accumulation of noise as time elapses (Kinchla & Smyzer, 1967; Lee & Harris, 1996). The other is deterministic (Gold, Murray, Sekuler, Bennett, & Sekuler, 2005), and implies that decay happens in a systematic and predictable fashion. According to the latter view, if a pattern was remembered and forgotten several times in a row, it would always be forgotten in a similar manner. One example of this would be to assume that visual memories are lost because they fade out over the course of time, meaning that they are always be forgotten in the same way –by means of blurring or contrast reduction. In stark contrast to the idea that memories decay gradually, stands the possibility that memories are lost in an all-or-none fashion. This option would imply that while some memory representations are maintained with a fixed resolution over an entire delay period, others are suddenly terminated at some point during their maintenance (Wei, Wang, & Wang, 2012; Zhang & Luck, 2009; Regan, 1985).

Whichever the mechanism, intuition dictates that memories are lost over time: With a stimulus gone from view it would be naïve to think it is maintained perfectly by the mind alone, without direct perception to keep it grounded. Nevertheless, ample psychophysical

evidence has suggested that memory resolution for a single simple visual features was limited only by its resolution during encoding, with little or no decay over time (Banko, Gal, & Vidnyanszky, 2009; Bennett & Cortese, 1996; Blake, Cepeda, & Hiris, 1997; Huang & Sekuler, 2010; Magnussen, Greenlee, Asplund & Dyrnes, 1991; Magnussen & Greenlee, 1992; Magnussen & Dyrnes, 1994; Magnussen, Greenlee, Aslaksen, & Kildebo, 2003; Regan, 1985; Regan & Beverley, 1985). Conversely, and in line with the intuition that visual memories cannot remain perfectly stable as time wears on, other evidence did show memory loss as a function of time (Fahle & Harris, 1992; Gold et al., 2005; Kinchla & Smyzer, 1967; Lee & Harris, 1996; Nilsson & Nelson, 1981; Phillips, 1974; Zhang & Luck, 2009). Such memory loss often suffers the sharpest falloff occurs earlier during retention rather than later (Fahle & Harris, 1992; Vogels & Orban, 1986) and has been described as exponential decay (Bisley & Pasternak, 2000; Lee & Harris, 1996; Nilsson & Nelson, 1981; Phillips, 1974). Exponential decay is common in nature, occurring for phenomena such as such as radioactive decay for large collections of atoms, or dissipating froth in a glass of beer (Leike, 2002).

Many of the above-mentioned studies, whether or not they found memory loss for simple visual features, have based their conclusions on change detection paradigms and threshold estimation. Such paradigms can be problematic as they might lack the sensitivity to detect memory loss over time (Skottun, 2004), and confuse the loss of precision with the loss of items from memory on a proportion of trials (Zhang & Luck, 2008). Moreover, discrimination thresholds determined via a two-alternative forced choice paradigm can readily be explained by the range of test stimuli employed, rather than by the sensory memory itself (Lages & Treisman, 1998). Recent evidence circumventing these possible confounds was obtained by utilizing a method-of-adjustment task, and uncovered memory loss in an all-or-none fashion (Zhang & Luck, 2009). Participants in this study remembered three colored squares (or shapes) over variable delay periods, after which they used a response wheel to indicate the color (or shape) most closely matching their memory contents. This method allowed for the construction of error distributions, to which a so-called 'mixture model' (Zhang & Luck, 2008) was fit. The 'mixture model'

approach summarizes memory errors into two classes: the variability with which items were maintained in memory (the SD of the error distribution) and the probability that items were completely forgotten and participants were guessing (the extent with which the error distribution was translated uniformly upwards). While this study found that guessing responses became more prevalent at longer delays, the amount of variability remained stable – favoring the notion that longer delays increased the likelihood of memory items suffering a 'sudden death', without any evidence to support gradual decay (Zhang & Luck, 2009).

Here we revisit the issue of temporal decay by examining the time course of visual memory for a variety of visual items: simple low-level features (orientation and color) as well as higher-level objects (faces). Participants remembered only a single item at a time, ensuring that they were operating well within the limits of their capacity. By utilizing a method-of-adjustment task we were able to take the entire distribution of responses to the memorized item into account – enabling a more in depth assessment of various aspects of visual memory and how such memories might endure the test of time. We found that visual short-term memories of orientation, color, and faces decayed gradually over time. Response variability increased with longer retention durations, while the probability of guesses only increased for a remembered color, but not for a remembered orientation or face.

Methods

Participants. Twelve healthy volunteers participated in each of the experiments. Initially, thirteen data sets were collected for Experiment 3, but one participant was excluded based on poor overall performance (> 2 SD from the group mean). Experiment 1 took place at Maastricht University under the approval of the standing ethical committee of the Psychology and Neuroscience department. Experiments 2 and 3 took place at Vanderbilt University under the approval of the Institutional Review Board of Vanderbilt

University. All participants at both locations reported normal or corrected-to-normal vision, and provided written informed consent. With the exception of authors RR (in Experiment 1) and YP (in Experiments 2 and 3) participants were naïve to the purpose of the study and received monetary reimbursement for their time. Participant's ages were between 21 and 36 (8 female) in Experiment 1, between XX and XX (X female) in experiment 2, and between XX and XX (X female) in experiment 3.

Stimuli. Participants viewed the stimuli in a dark room on a color and luminance-calibrated CRT monitor (Experiment 1: 1280 x 1024 resolution with 60 Hz refresh rate; Experiments 2 & 3: 1152 x 870 resolution with 75 Hz refresh rate). Visual stimuli were generated using MATLAB and the Psychophysics toolbox (Brainard, 1997; Pelli, 1997). Participants were seated at a viewing distance of 57 cm, and were instructed to maintain fixation throughout, aided by a chinrest and a central fixation bull's eye.

Stimuli in Experiment 1 consisted of centrally presented randomly oriented gratings (3° diameter; spatial frequency 2 c/deg; phase randomized; 20% Michelson contrast $\pm$ 10% uniform jitter). Gratings shared a mean luminance of 40.8 cd/m with the uniformly grey background and were presented within a wide Gaussian envelope (sd = 2.5°). Responses were obtained with a mouse-probe comprised of a white bull's eye (0.5° diameter) and an interrupted white line (each segment 0.025° wide and 0.125° long). The two lines were 3° apart in order not to overlap with the previous stimulus location. Moving the mouse made the dial rotate about fixation, allowing participants to replicate the memorized orientation.

Stimuli in Experiment 2 consisted of centrally circular colored patches (3° diameter), with the color randomly chosen from one of 360 color values evenly distributed along a circle in CIE L*a*b space (L = 70, a = 0, b = 0) with a radius of 45 units. Stimuli were presented against a grey background sharing the same luminance as the colored stimuli. By turning a knob (PowerMate 3.0, Griffin Technology, USA) participants could indicate their responses by moving a white circle around on a centrally presented color wheel (See

Figure 1A, middle panels). The color wheel was randomly rotated from trial to trial, as was the location of the white circle. Once a response was initiated, a centrally presented match-to-sample colored circle was added to the color wheel.

Stimuli in Experiment 3 consisted of grey-scale 3D face images (5.2° by 6.5°) generated with FaceGen Modeller software (Singular Inversions Inc.) as in Lorenc, Pratte, Angeloni, & Tong (2014), and presented against a black background. Face images were normalized to equate for mean luminance. Eight faces varying along dimensions of age and gender (Figure 1B), forming an octagonal space, were generated first. Next, each pair of neighboring faces were morphed together linearly in varying proportions (10/90, 20/80, … 90/10), ultimately resulting in a set of 80 unique faces which we consider to be spaced evenly along a 360° approximately circular 'face space'. This implies that for our analyses, each face in this space is 4.5° apart from its neighbor. During the experiment, faces were centered on a grey bull's eye fixation at eye height, around the nasal bridge.

Procedure. The general outline of the tasks during all experiments (Figure 1A) was that participants viewed a randomly chosen sample stimulus for 200ms (Experiments 1 and 2) or 500ms (Experiment 3), and they remembered this stimulus for a duration of 1, 3, 6, or 12 seconds (randomly interleaved). After the retention phase participants were presented with a test stimulus, and used a mouse (Experiment 1) or knob (Experiments 2 and 3) to provide an unspeeded response, replicating the stimulus in memory as precisely as possible. In Experiment 1 the interrupted line rotated about fixation by movements of the mouse. In Experiment 2 the knob was rotated to make the response dot move along the color wheel, and upon the first knob movement (after participants had the chance to view the color wheel and the location of the response dot on the wheel) a central color patch appeared that morphed through color space in response to rotating the knob. In Experiment 3 a centrally presented probe face morphed through face-space by turning the knob, allowing participants to arrive at their desired response. Once satisfied with their response, participants clicked the mouse or knob to continue.

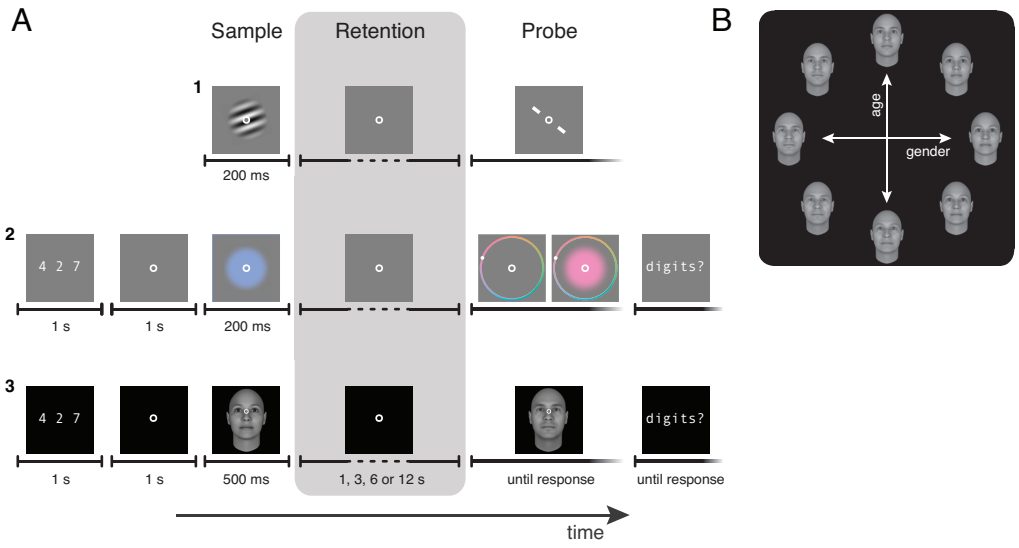


Figure 1. Trial sequence and stimuli. **(A)** In Experiment 1 (top row) participants viewed a randomly oriented sample orientation for 200ms at the start of each trial, and remembered it as precisely as possible for a duration of 1, 3, 6, or 12 seconds. After the retention interval participants were presented with a test stimulus, which they could rotate by using the computer mouse to match the memorized orientation. When satisfied with their response, participants clicked the mouse and continued to the next trial 1 second later. During Experiment 2 (middle row) participants first viewed three digits that they repeated aloud throughout the trial. After a short blank they viewed a colored circle for 200ms, remembering the color for 1, 3, 6, or 12 seconds. The memory probe was a randomly rotated color wheel with a white response circle, and once participants started turning the knob a color patch appeared which they matched to the color in memory. Finally, the three digits were entered on the keyboard. Experiment 3 (bottom row) was identical to Experiment 2 with the exception that stimuli consisted of 3D rendered face images presented against a black background. The probe appeared centrally, at the same location as the stimulus, and turning the knob made the face morph through face space in order to arrive at the desired response. **(B)** The eight originally generated samples that define the 'face space' along the dimensions of age and gender. For illustrative purposes the dimension of the stimuli above are not to scale – see the text for actual sizes.

For Experiment 1 memory stimuli were oriented gratings, which map naturally onto a circular space that is difficult to verbalize (Figure 1A top row). However, both colors (Figure 1A middle row) and faces (Figure 1A bottom row) lend themselves well to the use of verbal strategies, which is why we included a verbal suppression component to these two experiments: At the start of each trial participants were presented with three digits for

1 second, followed by a 1 second interval after which they performed the visual memory part of the trial. Participants repeated the three digits aloud until after they were done replicating the memorized color or face, at which point they used the keyboard to input the digits. After the final mouse or button press at the end of each trial, the time until the next trial (for all experiments) was one second. For Experiment 1 we collected 100 trials per condition, for Experiments 2 and 3 we collected 90 trials per condition.

Analyses. For each condition of interest a distribution of response errors was obtained by calculating the difference between memory target and response (reported stimulus minus sample stimulus). Memory accuracy is the average (absolute) error in degrees. Relevant characteristics from these response distributions were estimated by fitting two models that have been previously proposed to describe various aspects of the working memory system. The first is known as the mixed-model approach (Zhang & Luck, 2008), which assumes that memory contents can be described by two parameters: the standard deviation (SD) of a distribution which is believed to represent memory precision, and the extent to which the distribution needs to be translated uniformly upwards which is believed to represent the probability of memory failure (P-uniform). Furthermore, the distribution mean (μ) indicates the center of the distribution (presumably centered around 0° error), and was included to investigate potential response drifts. The second model used is a von Mises function (circular analog of a normal distribution), which describes the data only in terms of the mean (μ) and circular variance (SD).

Data analysis were performed in MATLAB using custom functions as well as functions provided by the Bays lab (Bays, Catalao, & Husain, 2009), and the Circular Statistics Toolbox (Berens, 2009). Specifically, we used Maximum Likelihood Estimation to perform fitting and estimate parameters (on which we perform regular repeated-measures statistics).

Results

Orientation. Error histograms and stimulus-response plots are shown in Figure 2. These data give the general impression that reports became noisier as a single orientation had to be retained over longer delays.

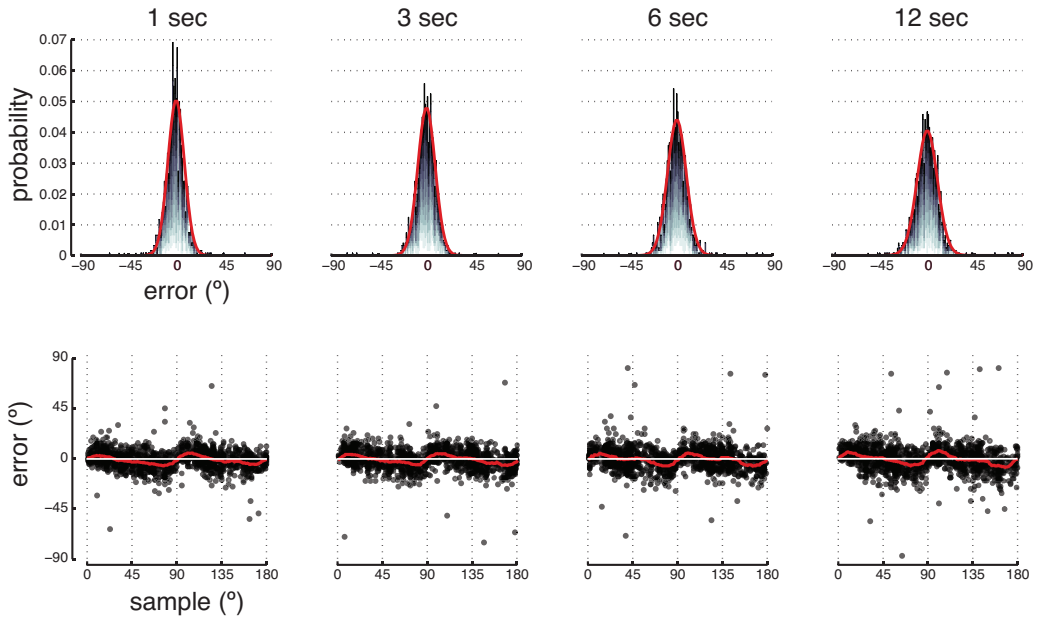


Figure 2. Orientation data for 1, 3, 6, and 12-second delay durations. Top panels show the error histograms with each individual participant depicted in a different shade of grey. Mixed-model fits are overlaid in red. Response errors became more variable as a single orientation was retained for longer, without an obvious increase in random responses. Bottom panels depict the same data by plotting the response errors against stimulus orientation. Red lines depict the mean error calculated over a sliding window. Participants' reports were biased away from cardinal orientations. Reports appeared generally noisier at longer delay durations.

To quantify what happened to memory representations over longer delays, we first calculated the circular variance (V, independent of the mean report) of the reported error in each delay condition (Figure 3, left most panel). The variance increased with longer retention durations ($F_{(3,33)} = 8.62$; $p < 0.001$). Paired t-tests (uncorrected) showed that this

difference was significant between delays of 1 and 6 seconds ($t_{(11)}$ = 3.9; p = 0.003), 1 and 12 seconds ($t_{(11)}$ = 3.44; p = 0.006), 3 and 6 seconds ($t_{(11)}$ = 2.73; p = 0.02), and 3 and 12 seconds ($t_{(11)}$ = 3.14; p = 0.009). The average time that participants took to respond during the various delay conditions is also shown in the left most panel of Figure 3 – response times were slower with longer delay durations ($F_{(3,33)}$ = 5.128; p = 0.005). Particularly, response times were longer when the delay was 12 seconds, compared to when the delay was 1 ($t_{(11)}$ = 2.7; p = 0.02), 3 ($t_{(11)}$ = 2.2; p = 0.05), or 6 ($t_{(11)}$ = 2.42; p = 0.03) seconds.

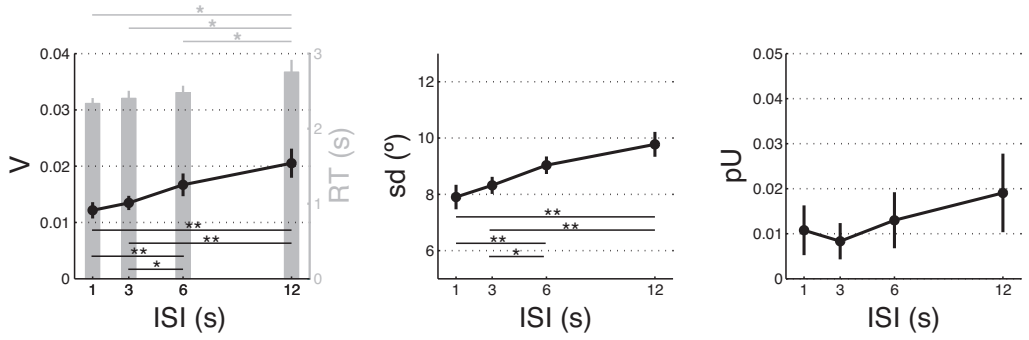


Figure 3. Main results for orientation memory. Left panel shows an increase in both response time and the circular response variability as a single orientation had to be maintained in memory longer. The middle panel shows an increase in the variability (calculated with a mixed-model approach) at longer delays, while the right panel shows no significant changes in the probability of random responses as a function of delay.

In order to investigate the contributions of an increase in report variability and the rate of forgetting, we summarized the error distributions (using a mixture-model) into a standard deviation (sd) and probability of uniform responses (pU) respectively. While response variability increased at longer delay durations ($F_{(3,33)}$ = 11.86; p < 0.001), the probability of uniform responses did not change as a function of delay ($F_{(3,33)}$ = 1.123; p = 0.354). Paired t-tests (uncorrected) demonstrated variability differences between delays of 1 and 6 seconds ($t_{(11)}$ = 4.03; p = 0.002), 1 and 12 seconds ($t_{(11)}$ = 4.43; p = 0.001), 3 and 6 seconds ($t_{(11)}$ = 2.34; p = 0.03), and 3 and 12 seconds ($t_{(11)}$ = 4.2; p = 0.002).

Color. As for the orientation data, error histograms and stimulus-response plots are displayed in Figure 4. While responses to a remembered orientation did not appear more random at longer delay durations, responses to a remembered color do appear random at times, and more so during longer delays.

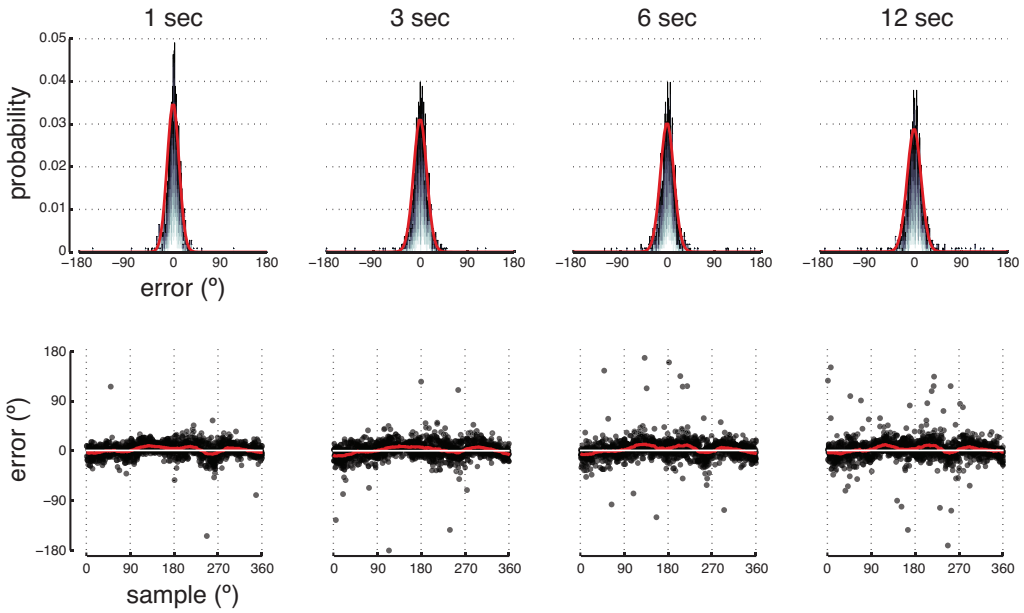


Figure 4. Color data for 1, 3, 6, and 12-second delay durations. Top panels show the error histograms with each individual participant depicted in a different shade of grey, in red the mixed-model fits across all participant's data are shown. Bottom panels plot the error in color report against the sample color. Participants' reports showed some systematic biases at all delay durations. In general, errors in report appeared noisier as a single color was retained for longer, with some increase in random responses.

The circular variance (V) increased at longer delays ($F_{(3,33)} = 6.88$; $p = 0.01$; left most panel of Figure 5), as did the response times ($F_{(3,33)} = 16.38$; $p < 0.001$; all paired t-test $p < 0.008$). Increases in circular variability were significant between delays of 1 and 3 ($t_{(11)} = 2.44$; $p = 0.03$), 6 ($t_{(11)} = 3.48$; $p = 0.005$) and 12 ($t_{(11)} = 3.68$; $p = 0.004$) seconds, as well as between delays of 3 and 12 seconds ($t_{(11)} = 2.35$; $p = 0.04$).

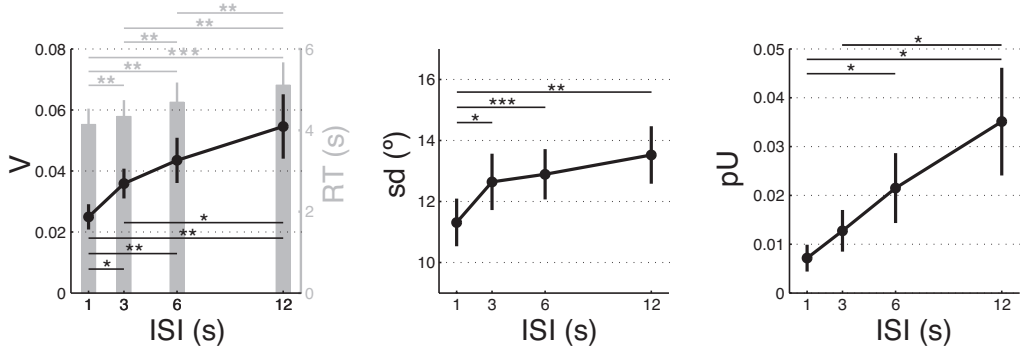


Figure 5. Main results for color memory. Left panel shows an increase in both response time and the circular response variability when a single colored patch was maintained in memory over a longer delay. Mixed model parameters showed an increase in the variability (middle panel) as well as an increase in the probability of random responses (right panel) as a function of delay.

The middle and right most panel of Figure 5 show that with longer retention delays there was an increase in variability ($F_{(3,33)} = 6.05$; $p = 0.002$) as well as the probability of guesses ($F_{(3,33)} = 4.83$; $p = 0.007$). Variability differences existed between the shortest 1-second delay and all other delays (3-second $t_{(11)} = 2.44$; $p = 0.03$, 6-second $t_{(11)} = 5.3$; $p < 0.001$, 12-second $t_{(11)} = 3.78$; $p = 0.003$). Random responses were more prevalent for delay durations of 12 seconds compared to 1 ($t_{(11)} = 2.91$; $p = 0.014$) and 3 ($t_{(11)} = 2.22$; $p = 0.048$) seconds, and for a delay of 6 compared to 1 seconds ($t_{(11)} = 2.38$; $p = 0.037$).

Faces. Error histograms and stimulus-response plots are shown in Figure 6. These data give the general impression that reports became noisier as a single orientation had to be retained over longer delays. Moreover, when a face had to be remembered (as opposed to a single orientation or color) reports appeared considerably noisier overall.

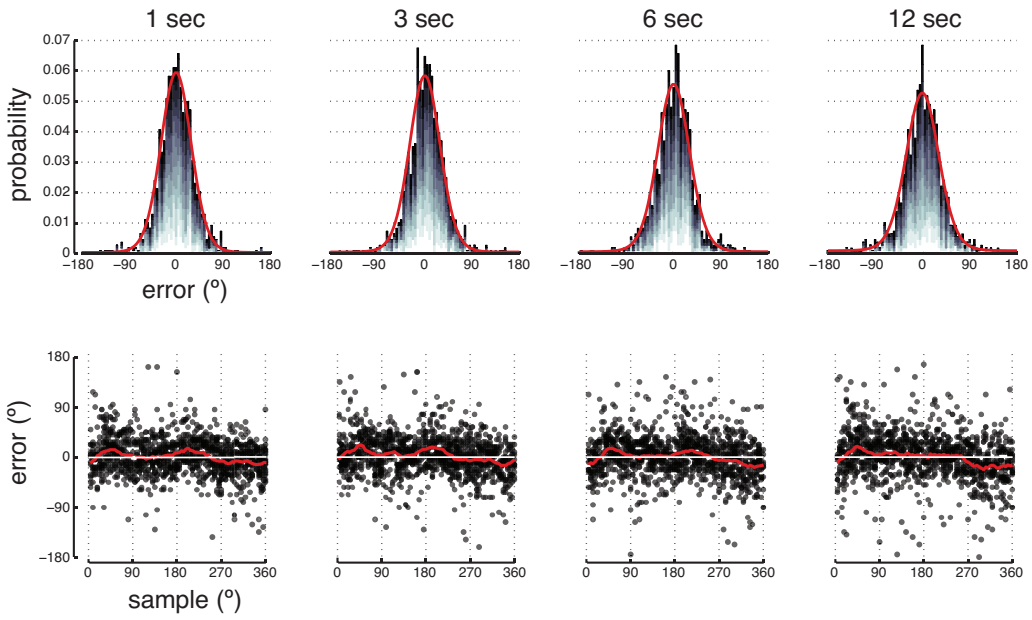


Figure 6. Memory for faces over 1, 3, 6, and 12-second delays. Top panels show the error histograms with each individual participant depicted in a different shade of grey, and mixed-model fits overlaid in red. Note that memory for a single face was much more variable than memory for orientation and color (compare histograms in this figure with those from Figures 2 and 4 respectively). Bottom panels plot the response errors against the sample face, demonstrating a clear increase in the noisiness of reports for longer delays. Systematic biases in face report are shown in red.

Similar to memory for orientation and color, response times increased when a face was remembered over a longer delay ($F_{(3,33)} = 15.95$; $p < 0.001$; all paired t-test $p < 0.03$, Figure 7 left panel). The circular variance also increased with longer retention durations ($F_{(3,33)} = 8.51$; $p < 0.001$), and this difference was most prevalent comparing 1 with 6 ($t_{(11)} = 2.34$; $p = 0.039$), 1 with 12 ($t_{(11)} = 4.62$; $p < 0.001$), 3 with 12 ($t_{(11)} = 3.3$; $p = 0.007$), and 6 with 12 ($t_{(11)} = 3.27$; $p = 0.007$) second delays.

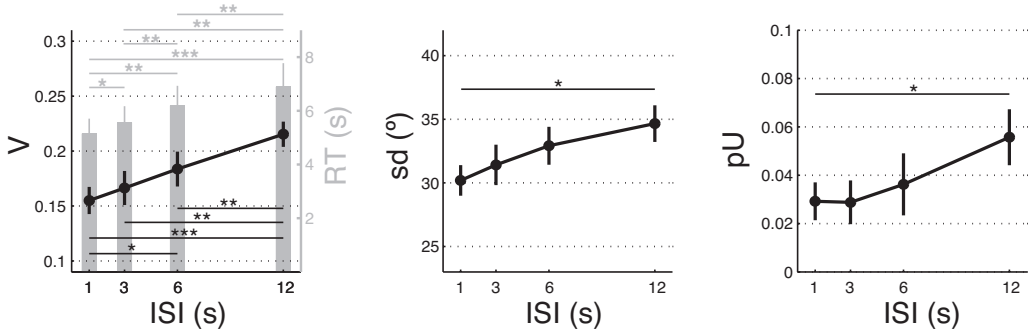


Figure 7. Memory for faces over various delay intervals. Left panel shows an increase in both response time and the circular response variability when a single face was maintained in memory longer. The mixed model standard deviation showed a trend indicating an increase in the variability (middle panel). While the ANOVA revealed no change in the probability of random responses (right panel), uncorrected paired t-tests indicated an increase in forgetting between 1 and 12 second delays.

Fitting the mixed model to the response distributions for face memory yielded a trend in the standard deviation, indicating an increase in response variability for longer delays ($F_{(3,33)} = 2.705$; $p = 0.061$). Uncorrected paired t-tests further suggest this difference exists, specifically, between delays of 1 and 12 seconds ($t_{(11)} = 2.68$; $p = 0.022$). The probability of guessing did not differ across the various delay intervals ($F_{(3,33)} = 1.77$; $p = 0.172$), however, uncorrected paired t-tests suggest more guessing for a 12 second compared with a 1 second delay ($t_{(11)} = 2.67$; $p = 0.022$).

Biases. To get a better impression of biases in report, we calculated the error at each stimulus orientation over a sliding window, collapsed across all participants. The size of the sliding window was standardized to yield ~10 observations per bin (i.e. window sizes of 19°, 41°, and 9° for orientation, color, and faces respectively). Biases calculated in this manner were already show in the bottom panels of Figures 2, 4, and 6 for orientation, color, and face memory respectively. Figure 8 replots these errors against the stimulus value at each delay (grey lines) and each experiment (panels).

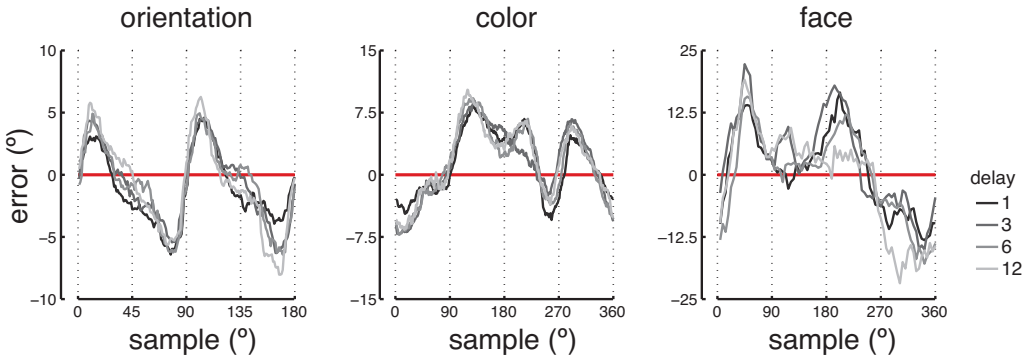


Figure 8. Memory biases in the report of orientation, color, and faces. For orientation, memory biases appear somewhat exaggerated as a function of delay interval, while for color and face memory any such differences are not obvious. Because here we were interested in directly comparing biased reports across retention intervals when participants were not guessing, we removed errors > 3 standard deviations before calculating the sliding window (note, such outlier removal was not performed for the red lines in the lower panels of Figures 2, 6, and 8).

Discussion

The representation of visual items in short-term memory – orientations, colors, and faces – became noisier with the passage of time. The probability of random responses did not change as a function of delay when an orientation or face was the memorized feature, while more guesses were observed when participants remembered a single color. First of all, these findings confirm that visual memories suffer representational loss as a function of time. Importantly, this was true for simple visual features (orientation, color), and extended to memory for complex objects (faces) as well. Secondly, these findings suggest that the main mechanism leading to memory loss over time is gradual decay due to the random accumulation of noise, under some circumstances accompanied by a sudden termination of representations. Furthermore, response times were increased at longer delays, with readily observable changes within the first 3 seconds for color and face memories, and between longer retention durations for orientation.

The finding that longer delays result in worse memory performance contradicts early psychophysical work showing little or no loss of simple visual features over time (Magnussen & Greenlee, 1999; Magnussen, 2000). However, much of this earlier work suffers from serious drawbacks. Some studies relied on very small sample sizes (Bennett & Cortese 1996; Magnussen & Greenlee, 1992; Regan, 1985; Regan & Beverley, 1985), and many investigated memory for spatial frequency and properties of motion (Blake et al., 1997; Huang & Sekuler, 2010; Magnussen et al., 1991; Magnussen & Greenlee, 1992; Regan, 1985; Regan & Beverley, 1985), which might involve storage processes separate from those of other visual features (Lages & Treisman, 1998; Lee & Harris, 1996; Magnussen & Greenlee, 1999; Pasternak & Greenlee, 2005). For example, it's been suggested that a distinction can be made between 'extensive' (such as spatial frequency, velocity, orientation, or hue) and 'intensive' (such as contrast or saturation) features: The first are coded along a distribution of activity in labeled detectors and are presumably not susceptible to memory loss, while the second are coded as the overall magnitude of activity and are susceptible to loss (Magnussen, Greenlee, & Thomas, 1996). A possible reason to favor the invariable maintenance 'extensive' but not 'intensive' features could be that the former usually remain constant over a period of observation, while the latter might vary under differences in illumination (Lages & Treisman, 1998). However, such an account does not hold for the present experiments encompassing stimuli that, by this dissociation, would be categorized 'extensive' while memory loss was clearly evident. Also, memory loss has been previously observed for other 'extensive' features, such as spatial memory measured with a vernier acuity task (Fahle & Harris, 1992), memory for location (Simmering, Peterson, Darling, Spencer, 2007), motion (Bisley & Pasternak, 2000), spatial frequency (Lages & Paul, 2006), or free-form figures (Cermak, 1971).

Furthermore, many early claims failing to show memory loss are not as conclusive as they appear at first glance. For example, one study testing memory for orientation over a 10-second delay found that thresholds did in fact rise as a function of time for both participants (4% and 37% differences respectively), but two-tailed t-tests performed separately for each of the two participants only revealed significance for one of them

(Regan & Beverley, 1985). Other work has claimed no loss of spatial frequency information over time when using a fixed sample stimulus despite statistical evidence of the contrary, and despite strong delay effects when using a wider range of sample stimuli (Magnussen et al., 1996). Still others have dismissed increased thresholds at longer intervals as being 'nearly constant' (Bennet & Cortese, 1996) stating they showed no 'appreciable decay' and attributing increases to 'lapses of attention' (Regan, 1985) or 'stimulus uncertainty' (Magnussen et al., 1996; Magnussen & Greenlee, 1999). Lapses of attention effectively equate to memory loss in an all-or-none manner, while stimulus uncertainty would imply that when one does not know the upcoming stimulus in advance memory retention is affected somehow.

How might uncertainty affect discrimination performance? Notably, method of constant stimuli paradigms with a fixed range of test stimuli centered on a single sample stimulus reliably produce thresholds representing the midpoint of the test stimuli (Lages & Treisman, 1998; Morgan, Watamaniuk, & McKee, 2000). This has been explained in a signal detection framework assuming fluctuations in the response criterion from trial to trial (Lages & Treisman, 1998), and imlies that findings of a good many studies aiming to investigate thresholds over time could be a mere artifact of test stimuli properties, rather than evidence of 'lossless' memory representations (Magnussen & Greenlee, 1992; Magnussen & Dyrnes, 1994). An attempt to bypass this confound has been made by calculating thresholds across large groups of individuals, each of whom only performed one or two trials, resulting in stable thresholds for a remembered spatial frequency (Magnussen et al., 2003) or face (Banko et al., 2009) over time. However, participants in these studies could have established an initial response criterion prior to the experiment by being allowed a practice with stimuli centered on the reference. Without such practice stimulus uncertainty is much higher, and thresholds were shown to rise considerably (Lages & Paul, 2006). Interestingly, the criterion-setting theory suggests that memory traces might not consist of sensory representations at all, but rather consist of specifications of the response criteria. In this view memory loss is a byproduct of the way in which criterion changes are set within the goal of optimizing performance, a strategy

that might generalize to other experimental paradigms as well (Lages & Treisman, 1998; Lages & Paul, 2006).

This is not to say that all failures to find memory loss as a function of time crumble under further scrutiny, but it does give a flavor of why discrepancies might have arisen in the literature regarding the effects of time on visual short-term memory. Besides the afore mentioned drawbacks, the amount of memory noise required to achieve observable changes in memory performance over time needs to be considerably large (Skottun, 2004). Such limits on the impact of time on visual memories might allow some delay related changes to be found while others are overlooked.

The way in which information is lost over time can prove critical to understanding the underlying mechanisms of memory maintenance. Do memories decay gradually over time, or do they terminate abruptly in an all-or-none fashion? In support of the latter, it has been proposed that cognitive states such as memory operate in a threshold manner, considered typical of biological systems driven by feedback (Zhang & Luck, 2009). It is indeed generally assumed that sensory representations are maintained in memory with the help of recurrent feedback (Durstewitz, Seamans, & Sejnowski, 2000; Pasternak & Greenlee, 2005; Sreenivasan, Curtis, & D'Esposito, 2014; Wang, 2001).

Mechanisms by which information might be maintained over time have been explored using dynamic neural network models, which exploit recurrent connections between neurons. Such models are complex but neurally plausible, generally consisting of a limited pool of resources shared by a distributed neural population with locally excitatory and laterally inhibitory connections. Constellations of connections between neurons can form an attractor state – a stable pattern of firing by which activity can be sustained over time. Arguments for both sudden death, as well as gradual decay have been made based on these models. In favor of gradual decay, network models for spatial working memory have demonstrated random drifts over time, or 'diffusion' (Brody, Romo, & Kepecs, 2003; Compte, Brunel, Goldman-Rakic, & Wang, 2000; Wang, 2001). Diffusion could account

for increasingly noisy behavioral responses as a function of time (Ploner, Gaymard, Rivaud, Agid, & Pierrot-Deseilligny, 1998; White, Sparks & Stanford, 1994; Wimmer, Nykamp, Constantinidis, & Compte, 2014). Besides diffusion and in support of sudden death, slow stochastic dynamics that unfold over time can result in the merging or fading out of analog memory representations, decreasing the probability that an item is represented in memory as time wears on. In other words, representations can indeed terminate suddenly under the influence of ongoing recurrent dynamic interactions (Simmering, Schutte, & Spencer, 2008; Wei et al., 2012), although this dependence on delay duration might only exist for intermediate 'critical' set sizes of about 3 to 4 items (Wei et al., 2012).

To empirically test gradual decay versus sudden death threshold measures are unsatisfactory, as they conflate the loss of precision with sudden termination of items on a proportion of trials. Instead, Zhang & Luck (2009) employed a method-of-adjustment task in combination with a mixed model approach to separate changes in variability (expected if memories decay gradually) from increased guesses over time (expected if memories terminate suddenly). They showed that guesses increased over time, while memory precision remained stable, which was taken as evidence for 'sudden death' of memory representations over longer intervals (Zhang & Luck, 2009). In contrast, here we demonstrated a loss of precision over time but no changes in guess rates when a single orientation or face was remembered, while using almost identical methods. By the same logic, our data support gradual decay over time instead of 'sudden death', directly contradicting previous efforts (Zhang & Luck, 2009). However, when our participants remembered a single color, increased guess rates were observed as a function of time alongside an increase in memory noise. This latter finding supports both the notion of gradual decay, as well sudden death of mnemonic representations, at least for color memory.

One notable difference with our study is that participants in the Zhang & Luck (2009) study remembered three items (a 'critical' set size according to neural network models)

rather than one, placing higher demands on attentional resources. However, this does not fully explain the discrepant findings, as the authors also mentioned a follow-up experiment with a single color presented on each trial, purportedly leading to the same pattern of results as with three mnemonic colors (Zhang & Luck 2009). Nevertheless, a distinction based on task difficulty might be an important one, as memory loss has been shown to depend on task difficulty, with faster loss over time for more difficult tasks (Laming & Sheiwiller, 1958; Pearson, Raškevičius, Bays, Pertzov, & Husain, 2014; Phillips, 1974; Posner & Konick, 1966), when complexity was increased along the task relevant dimension (Blake et al., 1997; Magnussen et al., 1996).

The evidence presented here strongly implicates that visual memories dwindle due to a random accumulation of noise over time. This is further supported by increased response times observed as a function of delay. Longer response times indicate that participants became unsure of their memories with the passage of time, which could happen if their representations became noisier (Nilsson & Nelson, 1981). Indeed, it's been shown that by assuming a decision process based on noisy evidence, the presence of more noise leads to slower decisions and longer response times (Pearson et al., 2014).

Gradual decay via the accumulation of random noise (Kinchla & Smyzer, 1967) is at odds with an alternative explanation, stating that memories decay gradually in a deterministic fashion. One example of deterministic decay is the low-pass filtering of visual memories, which effectively means that an image in memory literally fades away over time. One study directly comparing stochastic decay to deterministic decay by means of gradual fading, found that the point of subjective equality did not change when observers compared two gratings of different contrasts over increasingly longer delays, demonstrating that the gratings did not fade from memory. The just noticeable difference between the two contrasts did increase however, supporting the notion of a noise accumulation (Lee & Harris, 1996).

However, a low-pass filter hypothesis is not the only deterministic explanation that may account for memory decay. For example, memories might blur as a function of time due to the systematic removal of fine details (Gold et al., 2005). Alternatively, a change in the mean response at varying retention intervals could capture systematic drifts in a memory representation (Simmering, et al., 2007; Spencer & Hund, 2002). A third example would be memory representations converging onto a central stimulus value over time, like the average size (Brady & Alvarez, 2011) or spatial frequency (Huang & Sekuler, 2010; Wilken & Ma, 2004) of a stimulus set. Note that here we did not observe any changes in the mean response as a function of time (orientation $F_{(3,33)} = 0.712$; $p = 0.552$, color $F_{(3,33)} = 0.55$; $p = 0.652$, face $F_{(3,33)} = 0.55$; $p = 0.652$). Nor did representations converge onto a central value, the possibility of which was circumvented in the current experiments by the circular nature of our stimuli. Although we did observe biases in report (Figure 8), those did not appear to be subject to obvious systematic changes as a function of delay, which is in accordance with previous findings for direction of motion (Blake et al., 1997), location (White et al., 1994) and color (Nilsson & Nelson, 1981) biases.

A final caveat when interpreting the evidence presented here as favoring the random accumulation of noise over time ties into a frequently debated question regarding the fundamental properties of visual working memory system. One proposal is that items are stored in 3-4 discrete slots with a fixed resolution (Zhang & Luck, 2008; Luck & Vogel, 1997), while others propose a fixed resource that can be flexibly distributed amongst items without a limit on the number of remembered items (Bays, 2015; Bays & Husain, 2008; Bays et al., 2009; Fougnie et al., 2012; van den Berg, et al., 2012; Wilken & Ma, 2004). Biological underpinnings of a fixed resource could be achieved by the encoding of information in the combined activity of a pool of neurons, known as population coding (Bays, 2014; 2015). At first glance our findings seem to support a stochastic and variable memory resource, demonstrating flexibility in precision over time. However, we cannot completely rule out notion of slots. The slots-and-averaging account (Zhang & Luck, 2008) suggests that when a single item is remembered, each slot can be utilized to hold a noisy estimate of that item. When more slots are available, averaging across them will lead

to a more precise estimation. In our experiment representations in some slots may have been dropped on a subset of trials. If the likelihood of a slot being dropped increases as a function of time, this might cause a reduction in the precision of the report –only when all 3 slots are dropped will a random guess ensue. However, we do not believe this explanation is likely, and would speculate that similar noise increases are to be expected over time for set size three or larger also. Indeed, when four orientations had to be remembered together, decaying memory traces were not irretrievably lost; an attentional cue during retention nearly abolished temporal decay (Pertzov, Bays, Joseph, & Husain, 2013). Furthermore, error distributions obtained in our experiments demonstrated a strongly 'peaked' shape, a deviation from normal typically observed in working memory tasks (Bays 2014; Fougnie et al., 2012; van den Berg et al., 2012), and one that cannot be accounted for by the slots-and-averaging account (Bays, 2015).

In sum, we observed memory loss for simple features as well as complex objects via the accumulation of noise with the passage of time. Memory loss is likely an unavoidable feature of the brain –a closed system with finite resources. It is possible that memories terminate abruptly and completely from time to time, although the evidence presented here suggests that the primary cause of memory loss over the course of time is the gradual decay of visual memory traces. Gradual decay might serve a functional role, for example as a mechanism for representing the temporal order of memory items (Warden & Miller, 2007). Studying the limitations and utility of mechanisms by which memories are lost could prove valuable for understanding the memory system as a whole.

References

Banko, E.M., Gal, V., & Vidnyanszky, Z. (2009). Flawless visual short-term memory for facial emotional expressions. *Journal of Vision*, *9*(1), 12–12.

Bays, P.M. (2014). Noise in Neural Populations Accounts for Errors in Working Memory. *Journal of Neuroscience*, *34*(10), 3632–3645.

Bays, P.M. (2015). Spikes not slots: noise in neuralpopulations limits working memory. *Trends in Cognitive Sciences*, *19*(8), 431–438.

Bays, P.M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Bays, P.M., Catalao, R.F.G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11.

Bennett, P.J. & Cortese, F (1996). Masking of spatial frequency in visual memory depends on distal, not retinal, frequency. *Vision Research, 36*(2), 233–238.

Bisley, J.W., & Pasternak, T. (2000). The multiple roles of visual cortical areas MT/MST in remembering the direction of visual motion. *Cerebral Cortex*, *10*(11), 1053–1065.

Blake, R., Cepeda, N.J., & Hiris, E. (1997). Memory for visual motion. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(2), 353.

Brady, T.F., & Alvarez, G.A. (2011). Hierarchical encoding in visual working memory: ensemble statistics bias memory for individual items. *Psychological Science*, *22*(3), 384–392.

Brody, C.D., Romo, R., & Kepecs, A. (2003). Basic mechanisms for graded persistent activity: discrete attractors, continuous attractors, and dynamic representations. *Current Opinion in Neurobiology*, *13*(2), 204–211.

Cermak, G.W. (1971). Short-term recognition memory for complex free-form figures. *Psychonomic Science, 25*(4), 209–211.

Compte, A., Brunel, N., Goldman-Rakic, P.S., & Wang, X.J. (2000). Synaptic mechanisms and network dynamics underlying spatial working memory in a cortical network model. *Cerebral Cortex*, *10*(9), 910–923.

Durstewitz, D., Seamans, J.K., & Sejnowski, T.J. (2000). Neurocomputational models of working memory. *Nature Neuroscience Supplement*, *3*, 1184–1191.

Fahle, M. & Harris, J.P. (1992). Visual memory for vernier offsets. *Vision Research, 32*(6), 1033–1042.

Fougnie, D., Suchow, J.W., & Alvarez, G.A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229.

Gold, J.M., Murray, R.F., Sekuler, A.B., Bennett, P.J., & Sekuler, R. (2005). Visual memory decay is deterministic. *Psychological*, *16*(10), 769–774.

Huang, J., & Sekuler, R. (2010). Distortions in recall from visual memory: two classes of attractors at work. *Journal of Vision*, *10*(2):24, 1–27.

Kinchla, R.A., & Smyzer, F. (1967). A diffusion model of perceptual memory. *Perception & Psychophysics 2*(6), 219–229.

Lages, M., & Paul, A. (2006). Visual long-term memory for spatial frequency? *Psychonomic Bulletin & Review*, *13*(3), 486–492.

Lages, M., & Treisman, M. (1998). Spatial frequency discrimination: visual long-term memory or criterion setting? *Vision Research, 38*(4), 557–572.

Laming, D., & Scheiwiller, P. (1985). Retention in perceptual memory: a review of models and data. *Perception*

& Psychophysics, 37(3), 189–197.

Lee, B., & Harris, J. (1996). Contrast transfer characteristics of visual short-term memory. *Vision Research*, *36*(14), 2159–2166.

Leike, A. (2001). Demonstration of the exponential decay law using beer froth. *European Journal of Physics*, *23*, 21–26.

Lorenc, E.S., Pratte, M.S., Angeloni, C.F., & Tong, F. (2014). Expertise for upright faces improves the precision but not the capacity of visual working memory. *Attention, Perception, & Psychophysics*, *76*(7), 1975–1984.

Luck, S.J., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Magnussen, S., Greenlee, M.W., Asplund, R., & Dyrnes, S. (1991). Stimulus-specific mechanisms of visual short-term memory. *Vision Research*, *31*(7-8), 1213–1219.

Magnussen, S., & Greenlee, M.W. (1992). Retention and disruption of motion information in visual short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 151–156.

Magnussen, S., & Dyrnes, S. (1994). High-fidelity perceptual long-term memory. *Psychological Science, 5*(2), 99–102.

Magnussen, S., Greenlee, M.W., & Thomas, J.P. (1996). Parallel processing in visual short-term memory. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(1), 202–212.

Magnussen, S., & Greenlee, M.W. (1999). The psychophysics of perceptual memory. *Psychological Research*, *62*(2), 81–92.

Magnussen, S. (2000). Low-level memory processes in vision. *TRENDS in Neurosciences*, *23*(6), 247–251.

Magnussen, S., Greenlee, M.W., Aslaksen, P.M., & Kildebo, O.Ø. (2003). High-Fidelity Perceptual Long-Term Memory Revisited: And Confirmed. *Psychological Science*, *14*(1), 74–76.

Morgan, M.J., Watamaniuk, S.N., & McKee, S.P. (2000). The use of an implicit standard for measuring discrimination thresholds. *Vision Research*, *40*(17), 2341–2349.

Nilsson, T.H., & Nelson, T.M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 141–150.

Pasternak, T., & Greenlee, M.W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience*, *6*(2), 97–107.

Pearson, B., Raškevičius, J., Bays, P.M., Pertzov, Y., & Husain, M. (2014). Working memory retrieval as a decision process. *Journal of Vision*, *14*(2):2, 1–15.

Pertzov, Y., Bays, P.M., Joseph, S., & Husain, M. (2013). Rapid forgetting prevented by retrospective attention cues. *Journal of Experimental Psychology: Human Perception and Performance*, *39*(5), 1224–1231.

Phillips, W.A. (1974) On the distinction between sensory storage and short-term visual memory. *Perception & Psychophysics, 16*(2), 283–290.

Ploner, C.J., Gaymard, B., Rivaud, S., Agid, Y., & Pierrot-Deseilligny, C. (1998). Temporal limits of spatial working memory in humans. *The European Journal of Neuroscience*, *10*(2), 794–797.

Posner, M.I. & Konick, A.F. (1966). Short-term retention of visual and kinesthetic information. *Organizational Behavior and Human Performance, 1*, 71–86.

Regan, D. (1985). Storage of spatial-frequency information and spatial-frequency discrimination. *Journal of the Optical Society of America, 2*(4), 619–621.

Regan, D., & Beverley, K.I. (1985). Postadaptation orientation discrimination. *Journal of the Optical Society of America a, Optics and Image Science*, *2*(2), 147–155.

Simmering, V.R., Peterson, C., Darling, W., & Spencer, J.P. (2007). Location memory biases reveal the challenges of coordinating visual and kinesthetic reference frames. *Experimental Brain Research*, *184*(2), 165–178.

Simmering, V.R., Schutte, A. R., & Spencer, J.P. (2008). Generalizing the dynamic field theory of spatial cognition across real and developmental time scales. *Brain Research*, *1202*, 68–86.

Skottun, B.C. (2004). On the use of discrimination to assess memory. *Perception & Psychophysics*, *66*(7), 1202–1205.

Spencer, J.P., & Hund, A.M. (2002). Prototypes and particulars: Geometric and experience-dependent spatial categories. *Journal of Experimental Psychology General*, *131*(1), 16–37.

Sreenivasan, K.K., Curtis, C.E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*(2), 82–89.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785.

Vogels, R., & Orban, G.A. (1986). Decision processes in visual discrimination of line orientation. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 115–132.

Wang, X.J. (2001). Synaptic reverberation underlying mnemonic persistent activity. *TRENDS in Neurosciences*, *24*(8), 455–463.

Warden, M.R., & Miller, E.K. (2007). The Representation of Multiple Objects in Prefrontal Neuronal Delay Activity. *Cerebral Cortex*, *17*(suppl 1), i41–i50.

Wei, Z., Wang, X.J., & Wang, D.H. (2012). From Distributed Resources to Limited Slots in Multiple-Item Working Memory: A Spiking Network Model with Normalization. *The Journal of Neuroscience*, *32*(33), 11228–11240.

White, J.M., Sparks, D.L., & Stanford, T.R. (1994). Saccades to remembered target locations: An analysis of systematic and variable errors. *Vision Research, 34*(1), 79–92.

Wilken, P., & Ma, W. (2004). A detection theory account of change detection. *Journal of Vision*, *4*(12).

Wimmer, K., Nykamp, D.Q., Constantinidis, C., & Compte, A. (2014). Bump attractor dynamics in prefrontal cortex explains behavioral precision in spatial working memory. *Nature Neuroscience*, *17*(3), 431–439.

Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.

Zhang, W., & Luck, S.J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*(4), 423–428.

Chapter 5

Investigating topographically specific effects of TMS over early visual cortex during visual working memory

Manuscript in preparation:

Rademaker R.L., van de Ven, V.G., Tong, F., Sack A.T.

Investigating topographically specific effects of TMS over early visual cortex during visual working memory.

Abstract

Recent imaging studies have demonstrated that patterns of activity in early visual areas are predictive of stimulus properties actively maintained in visual working memory. Yet, the mechanisms by which sensory areas represent such information remain largely unknown. In this study, observers were instructed to remember the orientations of 4 briefly presented gratings, one in each quadrant of the visual field. A 10Hz TMS triplet was applied either directly at stimulus offset, or midway through a 2-second delay, targeting early visual cortex corresponding retinotopically to a sample item in the lower hemifield. Memory for one of the four gratings was probed at random, and participants reported the probed orientation via method of adjustment. Replication errors were smaller when the visual field location targeted by TMS overlapped with that of the cued memory item, compared to errors for stimuli probed diagonally to TMS. This implied topographic storage of orientation information in early visual cortex, and a memory-enhancing effect at the targeted location. Furthermore, early TMS pulses impaired performance at all four locations in a topographically unspecific manner. Next, the errors were fit empirically using a mixture model analysis to characterize memory precision and guess rates. Memory was more precise for items proximal to the pulse location, irrespective of pulse timing. The probability of guessing was larger for early pulses, regardless of their proximity to the pulse location. Thus, whereas TMS administered at the offset of the stimulus array might disrupt early-phase consolidation, TMS otherwise acts to boost the precise representation of an item, perhaps by increasing attentional resources at its retinotopic location.

Introduction

Humans sense the world in a highly visual fashion – the flow of information from the eyes gives rise to an ostensibly effortless and seamless picture of our external environment. Despite its apparent simplicity, visual perception requires the brain to form an ongoing internal representation of all the information we are perceiving and perceived just moments ago, even if that information can no longer be sensed directly. Working memory is central to performing this complex cognitive task, allowing relevant information to be kept online for further computation, and serving as an indispensible buffer for human thought. As such, the study of working memory provides a unique window into human cognitive functioning. Here, we investigated working memory for visual information and the role of early visual cortex during the maintenance of such information.

How might the brain meet the computational demands associated with the maintenance of information to which it no longer has access? The act of keeping visual memories online involves frontal (Ester, Sprague, & Serences, 2015; Riggall & Postle, 2012) and parietal (Christophel, Hebart, & Haynes, 2012; Todd & Marois, 2004; Xu & Chun, 2006) regions, as well as visual areas that were involved when the information was originally sensed (Emrich, Riggall, LaRocque, & Postle, 2013; Harrison & Tong, 2009; Pratte & Tong, 2014; Serences, Ester, Vogel, & Awh, 2009). The coordinated effort of higher-level and sensory components during working memory is believed to be flexible and goal dependent (Lee, Kravitz, & Baker, 2013), and the dominant view is that higher-level areas recruit sensory areas that are specialized in processing the sensory analogs of memory contents (Awh & Jonides, 2001; D'Esposito, 2009; Jonides, Lacey, & Nee, 2005).

It has been suggested that sensory recruitment during visual memory is achieved in a spatially global and non-retinotopic manner: While people remembered an orientation presented in the left visual field, this orientation was decodable from fMRI signals originating from both ipsi- and contralateral V1 (Ester, Serences, & Awh, 2009). However, this particular task did not require subjects to maintain the relevant feature-contents

bound to any specific stimulus location. Therefore, the lack of retinotopically specific recruitment could easily be viewed as a spread of feature-based attention (Serences & Boynton, 2007; Sneve, Sreenivasan, Alnæs, Endestad, & Magnussen, 2015; Treue & Martínez Trujillo, 1999). Conversely, memory for visual information does depend on retinotopically specific representations when stimulus location is relevant to a task, and the explicit binding of stimulus contents to a particular location is required to perform it. Location matters when, for example, remembering objects in a scene (Hollingworth, 2006; 2007), when two orientations were presented one in each hemifield (Pratte & Tong, 2014), or when location was made salient by a spatial transformation during memory (Zaksas, Bisley, & Pasternak, 2001; Pasternak & Zaksas, 2003).

To directly probe the causal role of sensory areas during the retention of visual stimuli, as well as the spatial extent of such recruitment, memory can be actively disrupted by means of Transcranial Magnetic Stimulation ('TMS'). Previous work with TMS has provided support for both the necessity of visual sensory recruitment (Silvanto & Soto, 2012), as well as the retinotopically specific maintenance of representations in these areas (van de Ven, Jacobs, & Sack, 2012; Silvanto & Cattaneo, 2010). However, TMS studies demonstrating that the spatial extent of memory representations is confined in a retinotopic manner suffer some drawbacks: Specificity was only found very early during retention, probably during encoding (van de Ven et al., 2012), or measured indirectly, via the qualitative judgment of phosphenes (Silvanto & Cattaneo, 2010). Thus, while these studies suggest that brain stimulation has the potential to interfere with visual memories at a sensory level of representation, it remains to be seen whether retinotopically specific effects on performance can be found when TMS is applied outside of the range of sensory encoding.

While the existence of sensory recruitment during visual working memory is well documented, the functional role of such recruitment is much less understood. In addition to the issue of retinotopic specificity, here we aimed to address this pivotal question: What kinds of computations might sensory areas perform during visual memory maintenance,

and what is their functional role during sensory recruitment? Given that sensory areas can represent information with a degree of precision not easily achieved by less specialized areas, we assumed that their role in memory maintenance would be to maintain high-precision representations.

Here, questions of specificity and functionality were addressed by applying TMS over occipital cortex while participants were memorizing four oriented gratings, presented one in each quadrant of the visual field (Figure 1A). By cuing memory based on spatial location, this task encouraged participants to encode and retain orientation information at the spatial locations at which they were presented, which helps bind object identity to spatial information (Woodman, Vogel, & Luck, 2012), allowing us to test for retinotopically specific recruitment. To probe the functional role of sensory areas during the maintenance of orientation information, we employed a novel combination of methodologies: triple pulse TMS combined with rigorous psychophysical testing using the method of adjustment, the collection of very many trials per participant, and fitting a mixture model (Zhang & Luck, 2008) to the error distributions. A mixture model characterizes memory errors as having two underlying sources: response variability and the probability of uniform responses (Zhang & Luck, 2008). We applied this model to evaluate the effects of TMS on memory precision and the likelihood of successful memory. Furthermore, pulses were applied at two different time intervals to check for potential differences between processes occurring at the tail end of encoding, and processes occurring well within the retention phase.

Our primary hypothesis was that TMS should have disruptive effects on memory performance and precision, and that such disruption would occur in a retinotopically specific manner. This hypothesis was inspired by work showing a link between reduced information contents in visual cortex, as indexed by classification performance, and reduced mnemonic resolution (Emrich et al., 2013). And we assumed that the insertion of random noise with TMS would reduce the amount of information available in visual areas. Nevertheless, a decline in performance is not necessarily a given. Alternatively, TMS

over a retinotopically specific location might act to boost activity locally, which might be equivalent to drawing spatially specific attentional resources to that particular location.

Here we found that TMS during memory retention improved performance in a retinotopically specific manner. While we could not fully exclude the potential influence of baseline differences at the four stimulus locations, such differences were not likely to drive the TMS effects. TMS induced memory improvements were due to a reduction in response variability, without a change in the proportion of random responses. TMS early during retention, at the tail end of encoding, resulted in a global non-retinotopic performance decrement by increasing the likelihood of random responses without a change in mnemonic resolution.

Methods

Participants. Participants were 8 students from Maastricht University (5 females; mean age = 25.13 (SE = 0.81) years). All had normal or corrected-to-normal vision, and provided written informed consent. Before the start of the experimental proceedings participants completed a medical screening based on published safety guidelines (Rossi et al., 2009) and were approved for inclusion by an independent medical supervisor. The study was approved by the medical ethics committee of the Maastricht University Medical Centre. With the exception of one of the authors, participants received monetary reimbursement for their time.

Overall study design. For the purpose of this study we combined functional and anatomical MRI with neuro-navigated TMS during a working memory task (Figure 1A). This approach allowed us to keep the TMS stimulation site constant across experimental sessions, based on individually localized visual cortical activity. Participants were tested for a total of 7-8 separate sessions. During the first (fMRI) session, anatomical and functional localizer data were obtained. The second session was scheduled to determine the exact TMS target points, as well as the TMS intensity, that would be kept constant

throughout the remainder of the experiment. During this session, participants also practiced the psychophysical working memory task for a total of 160 trials in order to familiarize themselves with the paradigm, before the man experiment.

The following 5-6 sessions were used to collect psychophysical data on the working memory task, while applying TMS over visual cortex by means of fMRI-guided neuro-navigation (Sack et al., 2009). Specifically, a TMS coil (real or sham) was placed over either the left or the right dorsal part of early visual cortex (V1/V2) – where two individual target points were predefined based on the functional localizer fMRI session. For half of our participants, a session consisted of 4 blocks of 80 working memory trials each, of which three blocks involved triple-pulse TMS stimulation at 10 Hz during the retention phase of every trial, while one block involved sham stimulation. Target hemisphere (left or right) and type of stimulation (real or sham) were counterbalanced over blocks, sessions, and participants. The other half of participants underwent the same procedure, with one exception: they performed the blocks of sham-stimulation in a separate session, several months after completion of the real TMS sessions. This was done because randomly interleaving sham-blocks with blocks of real stimulation results in easily and directly observable differences in (tactile) sensations. Such differences could influence participant's behavior by for example introducing demand characteristics, or by systematically changing the manner in which attentional resources are allocated. Administering sham in two different ways allowed us check for such potential biases. The order in which blocks of sham TMS were administered did not influence participants overall performance (two-sample t-test: $t(6) = 1.012$, $p = 0.35$).

MRI measurements.

MRI acquisition. Scanning was performed at the Maastricht Brain Imaging Center (M-BIC) on a 3.0-Tesla Siemens MAGNETOM Allegra scanner using a standard birdcage head coil. A high-resolution 3D anatomical T1-weighted scan was acquired from each participant (FOV 256 x 256, 1 x 1 x 1 mm^3 resolution, 192 slices) by means of a

magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence. To measure BOLD contrast, standard gradient-echo echoplanar T2*-weighted imaging was used to collect 28 slices, which covered the entire occipital lobe as well as the posterior parietal and temporal cortex. For participants RR, LH, IB, TE, TN, and MS we used the following scan parameters: TR, 2000 ms; TE, 30 ms; flip angle, 80°; FOV 192 x 192; slice thickness, 3 mm (no gap); in-plane resolution, 3 x 3 mm^2. For participants AH and SR scan parameters were: TR, 2000 ms; TE, 30 ms; flip angle, 90°; FOV 256 x 256; slice thickness, 2 mm (no gap); in-plane resolution, 2 x 2 mm^2.

MRI data analysis. Preprocessing and analysis of the anatomical and functional MRI data were performed using BrainVoyager QX software (version 2.3.0.1750, Brain Innovation, Maastricht, the Netherlands). All anatomical data underwent inhomogeneity correction of signal intensity across space, and a tissue contrast enhancement using a sigma filter (7 cycles, range 5). Automatic grey-white matter segmentation was performed, after which manual corrections were made to improve the segmentation over the entire occipital cortex. The borders of the two resulting segmented sub-volumes were tessellated to produce surface reconstructions (folded meshes) – one for each hemisphere. These surface reconstructions were performed in order to recover the exact spatial structure of the cortical sheet and to improve the visualization of the anatomical gyrification.

functional MRI. Localizer stimuli in the scanner were generated using MATLAB 7.10.0 (R2010a) and Psychophysics Toolbox (Brainard, 1997). Stimuli consisted of 5 Hz flickering black-and-white checkerboards with a radius of 1° of visual angle, presented at 4° from fixation in either the lower left or lower right (randomly interleaved) quadrant of the screen against a uniform grey background (mean luminance of 55.86 cd/m^2). This stimulus location encompasses the same visual field position as the two lower Gabor patches in the Working Memory task (Figure 1A). A white bull's eye with a 0.5° radius was presented throughout to help participants maintain stable fixation. Stimuli were viewed through a mirror system on a back-projected screen with 1024 x 768 resolution and a 60 Hz refresh rate, at a distance of 66 cm in an otherwise darkened scanner room. We

presented our localizer in two separate functional runs, consisting of 12 blocks per run. Each block started with a 12 second fixation period, followed by a 12 second stimulus period during which the flickering checkerboard was presented. To ensure participants remained attentive, there was an occasional faint dimming of the checkerboard stimulus (~5 times per stimulus block), to which participants responded by means of a button press. This dimming was detected 43.18% of the time (SE=0.04). Each functional run was concluded by a final 12-second fixation period. In total, a run lasted 5 minutes.

Analysis of the functional localizer data was performed after discarding the first 4 volumes, after which automated 3D motion correction was applied. We also performed a slice-scan time correction (sinc) to correct for the different times of slice acquisition, as well as high pass temporal filtering (using GLM-Fourier basis set with 2 cycles) to correct for slow temporal drifts in the signal intensity of the data. No spatial or temporal smoothing was directly applied to the data. fMRI data was aligned to the within-session 3D anatomical scan by means of rigid-body transformations, after which all automated alignment was subjected to careful visual inspection and manual fine-tuning to correct for any residual misalignment. Volume time courses were created for both runs using sinc interpolation. Functional data obtained from both localizer runs were combined, and analyzed using a general linear model (GLM; Friston et al., 1995). We defined the functional target locations for both the left and right hemisphere based on the statistical contrast between the BOLD activities (Boynton, Engel, Glover, & Heeger, 1996) elicited by visual stimulation presented in the lower-right versus lower-left visual field quadrant respectively. Resulting functional activity maps were superimposed on the surface reconstructions of each participant's two hemispheres (the folded meshes).

Localization of the TMS target points. Positioning the TMS coil over the head was done using the BrainVoyager TMS Neuronavigation system (Brain Innovation, Maastricht, the Netherlands). Neuronavigation enables co-registration of an anatomical reconstruction of a participant's head with the participant's head in real space, by means of stereotaxic data recorded with an ultrasound digitizer. Miniature ultrasound senders

placed on the participant's head and on the coil transmit ultrasonic pulses to a receiving sensor, which determines spatial position in real 3D space based on the time these pulses take to travel. A local spatial coordinate system is defined by linking the raw position of the ultrasound senders in space to a fixed set of anatomical landmarks on the participant's head (nasion and the two tragi of the external ear), and a set of predefined landmarks on the TMS coil. Landmarks are specified in real space using a digitizer pen with two ultrasound senders. This procedure results in topographic information of the head and TMS coil relative to the ultrasound senders. Co-registration of real 3D space to MR space is achieved by specifying the same head and coil landmarks on the mesh of a participant's head, and on a digital version of the TMS coil respectively. After co-registration, movement of the TMS coil relative to the head in real space is visualized in real-time on a computer screen. This allows the TMS coil to be navigated to a position on the skull directly over a desired target region on the anatomical surface reconstruction of a participant's brain – such as an area of specific functional activation. Moreover, the system estimates the distance between the center of the TMS coil and the desired target on the cortical surface (coil-target distance), as well as the offset between the point of entry of the putative magnetic pulse 'beam' and the desired target (beam-target distance). This allowed us to position the TMS coil such that these distances were minimized. The use of fMRI-localizer-guided neuronavigation thus maximized the probability that the TMS pulses we administered primarily targeted the desired locations predefined during fMRI.

In the second experimental session (dedicated to determining individual TMS target points and TMS intensities) neuronavigation was used to manually maneuver the TMS coil relative to a participant's anatomical surface reconstruction on which the functional localizer activity was superimposed. The TMS target points, which would remain unchanged for the remainder of the experiment, were defined to always lie within this region of activation. More explicitly, each target point was determined to lie as posteriorly as possible within this region, while still eliciting a phosphene that overlapped with the location in the visual field where the stimuli would be presented (3-5° from fixation in either the lower left or lower right quadrant). After determining this location, each TMS

target point was indicated on the cortical surface by a digital marker, to guide future neuronavigation efforts. To conclude the second experimental session, individual phosphene thresholds for the left and right hemispheres were determined at the sites of these TMS target points (see TMS protocol), and were kept constant throughout the remainder of the experiment. For participants who did not experience phosphenes (TE and MS) we chose each target point to lie proximate to the peak-activity determined with fMRI, and the stimulation intensity to be the average intensity of the other participants in the study.

TMS protocol. During the working memory task we applied 10 Hz triple-pulse TMS. Biphasic TMS pulses were delivered by means of a figure-of-eight coil (MCB70) and a MagPro R30 stimulator (Medtronic Functional Diagnostics A/S, Skovlunde, Denmark; maximum stimulator output 1.9T). Pulses were applied at 80% of phosphene threshold to ensure that any TMS induced effects in our data could not be ascribed to simple visual interference caused by the appearance of phosphenes. Phosphene thresholds were individually determined for each participant during the second experimental session. Thresholds were obtained independently for the left and right hemisphere (evoking phosphenes approximately 50% of the time in the lower-right and lower-left visual quadrants respectively).

Participants received 240 TMS pulses (3 pulses * 80 trials) during each psychophysical run, and they performed a total of 16 experimental runs during which TMS was applied. In addition to the triplets during experimental runs, a variable amount of single pulses were applied during the second – localization and thresholding – session of the experiment. The average pulse intensity used in this experiment was 34.438% (SE=0.538) of maximum stimulator output. There were no significant differences between stimulation intensities for the two hemispheres (mean left = 35 %, mean right = 33.88%, $t = 1.386$, $p = 0.208$). The TMS coil was individually positioned such that we were able to elicit the clearest possible phosphenes, which aided in localization and threshold estimation. This resulted in coil orientations that were more or less locally parallel to the nearest sulcus

(relative to the target point) in some participants (AH & IB) and more or less locally perpendicular in others (RR, SR, LH, TE, TN, MS) including those participants who did not see phosphenes.

Working memory task.

Stimuli. All experimental stimuli were viewed in a dark room on a luminance-calibrated 19" Dell TFT monitor with 1280 x 1024 resolution and 60Hz refresh rate. Visual stimuli were generated with MATLAB 7.10.0 (R2010a) and the Psychophysics toolbox (Brainard, 1997) on a pc running Windows XP. Communication between the experimental pc and the stimulator was established by using PortTalk V2.0 (Beyond Logic). Stimuli consisted of randomly oriented gratings with a spatial frequency of 2 cycles/degree, extending 2° of visual angle in diameter, presented at four fixed locations around a central fixation point at an eccentricity of 4°. Gratings were presented at 20% Michelson contrast with a wide Gaussian envelope (sd = 2°) on a uniform grey background that shared the same mean luminance of 40.23 cd/m^2. Participants were seated at a viewing distance of 57 cm, and were instructed to maintain steady fixation throughout all experimental trials, aided by a centrally presented white bull's eye (0.5° of visual angle in diameter). A chinrest, in combination with tight placement of the TMS coil against the back of the head, assisted in maintaining head stability.

Procedure. Observers were presented with a 200 ms sample array consisting of 4 gratings to be retained in working memory (Figure 1A). Each grating had an independently chosen random orientation (1-180°), with the only constraint that orientations differed by at least 10°. During a 2-second retention interval a TMS triple-pulse was applied at 10 Hz (thus lasting 300 ms) either directly following the offset of the sample array, or midway through the retention interval (the first pulse occurring 900 ms into the interval). Next, 4 spatial cues appeared for 500 ms outlining the locations of the previously presented sample stimuli. A 0.15° thick white circle indicated the location of the target grating, which was to be reported from memory. Non-target locations were indicated by thinner outlines of

0.05°. Target location on a given trial was determined in a randomized fashion within each block of trials. After a brief 200 ms blank period, a test grating was presented centrally at an initially random orientation in order to probe participants' memory for the orientation of the cued target grating. Participants used separate buttons on a keyboard to rotate the test grating clockwise or counterclockwise to match the orientation in memory.

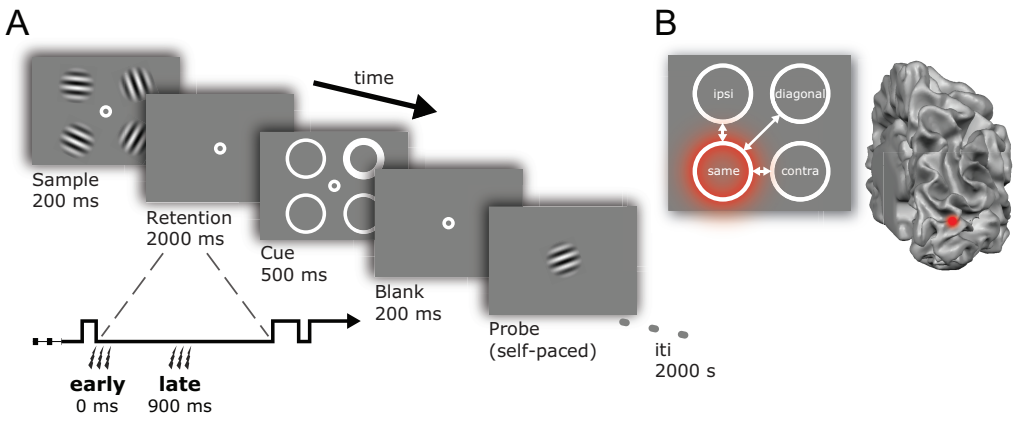


Figure 1. Trial sequence relative locations. **(A)** Participants view a sample array displaying 4 randomly chosen orientations, and remember these stimuli over a 2s interval. During the retention interval participants receive 3 pulses of real or sham TMS over their left or right hemisphere. The pulses arrive either directly at the offset of the sample array, or midway during the retention interval. A cue array indicates which of the four orientations is probed for recall, and after a short blank, participants rotate a test grating via button presses to match the orientation in memory. **(B)** Responses at the four visual field locations are analyzed according to their position relative to that of the pulse. Thus, the visual field position affected by the pulse might overlap with the memory item that is probed ('same'), the probed item might be contralateral to the affected visual field location ('contra'), it might be ipsilateral to it ('ipsi'), or diagonal to it ('diagonal'). In the example depicted here, the dorsal part of visual cortex in the right hemisphere is stimulated, resulting in disturbance at the lower-left visual field position. This defines the upper left position as 'ipsi', the upper right position as 'diagonal', the lower left as 'same', and the lower right position as 'contra' relative to the visual field location affected by the TMS pulse.

Analyses.

Relative location. Due to the anatomy of the human brain and head, TMS can only be applied over the dorsal (and not ventral) part of visual cortex. Thus, we were able to apply

TMS over only two of the four stimulus-locations probed during our memory task (in the lower-left and lower-right quadrants). The memory target can be cued with equal likelihood in all four quadrants. We analyzed our data looking at the relationship between the visual field location where the TMS pulse was applied, and the visual field location where the memory target was presented. Thus, we analyzed our data based on "relative location". To illustrate this, let's assume that for a particular experimental run we were stimulating the right hemisphere, hence targeting the lower-left visual field quadrant during the retention interval. If the target was subsequently cued in the same lower-left quadrant, both TMS and stimulus were presented at the "same" relative location. On trials where the target was cued in the upper-left quadrant (ipsilateral to the TMS-pulse location) we defined the relative location as "ipsilateral". Similarly, a target cued in the lower-right quadrant is at a "contralateral" relative location, whereas a target cued in the upper-right quadrant is at a "diagonal" relative location. The same logic can be applied when the coil is moved to the left hemisphere, thus targeting the lower-right visual quadrant. Data in the analysis are collapsed over both TMS sites.

Model fitting. To separately estimate the precision of memory for successfully remembered items and the likelihood of memory failure we adopted a mixture-model approach following the work of Zhang and Luck (2008). This model summarizes data from method-of-adjustment tasks in a way that reflects the underlying assumptions of the model: on some trials items are remembered with a certain degree of precision, whereas on other trials items are forgotten resulting in random guesses. This idea was implemented by fitting a circular Gaussian-shaped model to the distribution of orientation errors (actual orientation minus reported orientation) for each condition of interest. The model consisted of two key parameters: the standard deviation or width of the Gaussian distribution, and the extent to which the entire distribution needed to be translated along the y-axis to account for the frequency of uniform responses.

The mixed model assumes that the relative proportion of area under the curve corresponding to the uniform distribution reflects the probability of memory failure,

whereas the standard deviation of the error distribution reflects the precision of working memory for successfully remembered items. We rely on these summary statistics throughout this paper because they provide a useful way to summarize broad trends in the data and because they may signify distinct types of errors. However, it is important to acknowledge that the mapping between these summary statistics and underlying sources of error in the working memory system rely on assumptions regarding the exact nature of working memory performance, and that competing models have been proposed (e.g., Bays & Husain, 2008; Bays, Catalao, & Husain, 2009; Bays, 2014; Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012).

Results

Absolute performance. Experimental manipulations consisted of (1) probing memory for items at various relative distances from the visual field location targeted by TMS, and (2) applying TMS early or late during memory retention. Moreover, TMS was applied in a counterbalanced fashion either over the left- or right hemisphere, which was not expected to impact performance. By performing a 3-way within-subjects ANOVA (4 relative locations x 2 pulse timings x 2 stimulated hemispheres) on the absolute errors (absolute difference between reported and true orientation) we confirmed that the stimulated hemisphere (left or right) did not affect memory accuracy ($F_{(1,7)} < 0.001$; $p = 0.986$).

Proximity of a memory item to the visual field location targeted by TMS while items were being maintained in memory had a facilitative effect on memory performance, as indexed by smaller errors for items proximal to the pulsed location ($F_{(3,21)} = 3.951$; $p = 0.022$; Figure 2). Post-hoc ANOVA's showed that this effect was mainly due to the difference between trials on which the TMS pulses and probed location overlapped ('same' condition) versus when they were furthest apart ('diagonal' condition) ($F_{(1,7)} = 5.598$; $p = 0.050$). Post-hoc paired t-tests showed that when pulses were administered midway through the delay, performance was better at the targeted location compared to ipsilateral

($t_{(7)}$ = 2.402; p = 0.047) and diagonal ($t_{(7)}$ = 2.418; p = 0.046) locations. Post-hoc tests comparing TMS and sham trials show that performance was negatively affected for items probed diagonally relative to the pulses ($t_{(7)}$ = 2.762; p = 0.028), while all other direct comparisons with sham failed to reach significance (all p > 0.084).

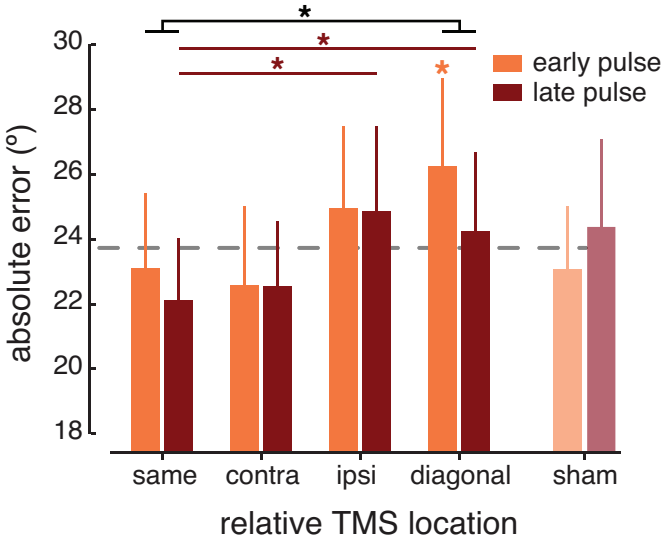


Figure 2. Absolute working memory performance at various visual field positions relative to the TMS pulses. When a memory item was probed at the same location as targeted by the TMS pulses during the retention interval, the absolute response error was smaller than when a memory item was probed diagonally relative to the pulses (black line and asterisk). For pulses midway through the delay, performance was better at the targeted location compared to locations ipsilateral and diagonal to the pulses (red lines and asterisks). Together, these results indicate better performance on trials where the probed item was more proximal to the site targeted with TMS. Note that the only significant difference from sham, however, occurred when pulses midway through the delay were administered diagonally to the location of the probed item (orange asterisk). In this case, performance was worse compared to sham. Early pulses resulted in worse performance than late pulses, irrespective of an item's location relative to the TMS pulses. Trails during which TMS was administered are shown here after collapsing across both hemispheres, since stimulation site (left or right hemisphere) did not affect participant's performance. Sham data is shown collapsed across all conditions. Error bars depict ± 1 SEM.

Comparing the two pulse-timings showed that applying TMS pulses early, directly at the offset of the stimulus display, resulted in larger errors than TMS applied midway through

the retention interval ($F_{(1,7)}$ = 6.359; p = 0.040), irrespective of location. It should be noted that when the data were collapsed across both hemispheres (as depicted in Figure 2) and the ANOVA was repeated with 2 factors (pulse timing and relative location) the latter finding was only marginally significant ($F_{(1,7)}$ = 4.820; p = 0.064), while performance differences between the four visual field locations remained statistically reliable ($F_{(3,21)}$ = 3.483; p = 0.034).

For data collected during the application of sham TMS we performed the same analyses as described above for real TMS, including as factors pulse timing, (absolute) location, and hemisphere. We did not find any significant differences in participant's performance across any of these conditions (all p > 0.265). Because sham TMS yields no neural effects, only pulse timing was considered as a condition with the potential to influence behavior, which is why a total of 320 sham trials (160 per pulse timing) were collected. This is much less than the 1280 trials collected during real TMS, and means that the ANOVA reported here may lack some statistical power as it is based on only ~20 trials per participant per condition (compared to ~80 for real TMS). Thus, we repeated the analysis of the sham data by first collapsing over the two hemispheres (resulting in ~40 trials per condition), again observing no significant results (all p > 0.323). Next, we separately looked at each factor (absolute location, pulse timing, and hemisphere) while collapsing across all the other factors, and again found no differences ($F_{(3,21)}$ = 0.459; p = 0.714 and $t_{(7)}$ = 0.647; p = 0.538 and $t_{(7)}$ = 1.013; p = 0.345 respectively). These outcomes are important, since some of our manipulations could have had unintended differential attentional-cuing effects, mimicking the neural effects probed via TMS. Considering that there were no differences between any of the conditions on sham trials – and the absence of a neural substrate associated with sham – we showed the sham data collapsed across all factors (dashed grey line in Figure 2), in addition to the sham data for each pulse timing condition (plotted in transparent colors on the far right of Figure 2).

Ultimately, we were interested in the question how TMS influenced memory for orientation, while being agnostic about the potential spread of TMS effects across the

visual field. To answer this question we calculated a difference metric by subtracting participant's performance at the TMS site from participant's performance averaged across non-TMS sites. This contrast is depicted in Figure 3A. Here, a positive difference indicates better performance when the probed item and TMS location overlapped ('same' condition), compared to performance at all other locations. Better performance at the TMS site reached significance for 'late' TMS trials during which the pulses were presented midway through the delay (two-sided $t_{(7)} = 2.387$; $p = 0.048$).

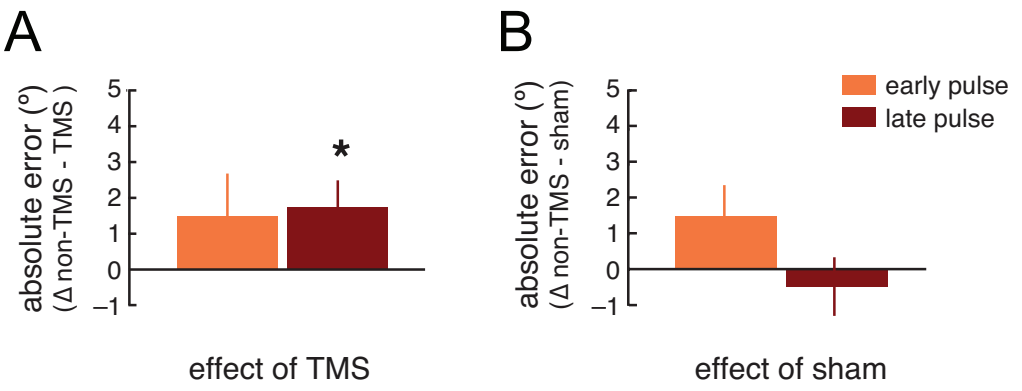


Figure 3. Difference metric for the absolute error. **(A)** Difference between absolute performance at the TMS-targeted location versus all other visual field locations. The positive difference score indicated that performance was better when the TMS pulses overlapped with the memory item, but only when the TMS pulses were applied midway through the retention interval. **(B)** Two conditions without an expected neural substrate were compared by subtracting performance on sham trials from performance on TMS trials during which stimuli were probed at locations not targeted by TMS. A positive difference score would indicate worse performance at non-TMSed locations compared to sham. However, as expected there was no difference between sham performance and performance at non-TMSed locations during real TMS trials. Error bars depict ± 1 SEM.

In the same vain, performance at the non-TMS sites (during TMS trials) was contrasted with performance on sham trials. Given that both sham and non-TMS sites can be considered control conditions without a neural substrate, this difference metric was expected to yield a score no different from zero. Figure 3B confirms that no performance differences existed between sham trials and TMS trials with the memory item probed at

one of the non-TMS locations (both two-sided p > 0.138). Together, the difference scores in Figure 3 give an indication of the neural effect due to TMS, with triple pulse TMS during retention resulting in improved performance at the location targeted by TMS.

Mixture-model results. To gain a deeper understanding into early visual cortex involvement during the maintenance of visual memories, we fit a so-called 'mixture model' to the data allowing us to decompose the absolute errors into two summary statistics: the precision of a memory representation (represented by the circular SD), and the likelihood of uniform 'guessing' responses (represented by the vertical displacement of the uniform portion of responses under the curve) (Zhang & Luck, 2008).

When a memory target was probed at a location proximal to the TMS pulses, memory recall was more precise ($F_{(3,21)}$ = 4.102; p = 0.019; Figure 4A). Specifically, post-hoc ANOVA's showed that memory was more precise on trials where the pulses and probed memory item either overlapped ($F_{(1,7)}$ = 7.974 ; p = 0.026) or were ipsilateral ($F_{(1,7)}$ = 7.658 ; p = 0.028) from one another, compared to trials on which they were furthest apart ('diagonal'). Pulse timing had no effect on memory precision ($F_{(1,7)}$ = 0.866; p = 0.383). Post-hoc paired t-tests showed that, for early pulses, precision was better for items probed at the 'same' compared to the 'contra' lateral ($t_{(7)}$ = 2.784; p = 0.027) location. Also, when the probed item overlapped with the location targeted by TMS, precision was higher than during sham ($t_{(7)}$ = 2.405; p = 0.047).

No retinotopic specificity was found for the probability of uniform responses, which did not differ between the four visual field locations ($F_{(1,7)}$ = 1.831; p = 0.172). However, participants were more likely to forget the orientation in memory when pulses were presented early during the retention interval compared to pulses presented midway through retention ($F_{(1,7)}$ = 6.594; p = 0.037; Figure 4B). This increase in random responses occurred irrespective of the location at which the memory item was probed (no interaction $F_{(3,21)}$ = 0.712; p = 0.555). Post-hoc comparisons showed that the difference in random responses between early and late pulses reached significance ipsilateral to the

pulses ($t_{(7)}$ = 3.013; p = 0.02), with more guesses for early pulses at this location compared to sham ($t_{(7)}$ = 2.77; p = 0.028).

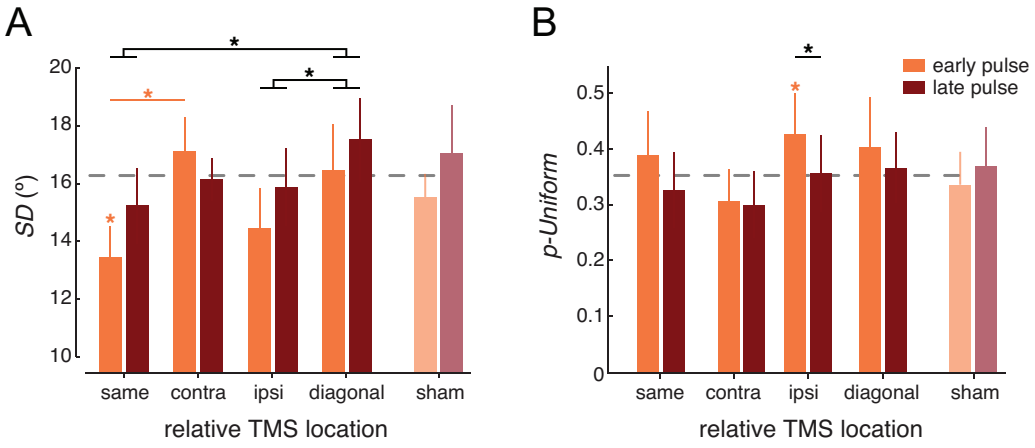


Figure 4. Model fits of the TMS data. **(A)** Memory precision is represented by the circular *SD*, with a smaller *SD* indicating more precision. Memory was most precise when the location of a memorized item overlapped with the location at which TMS was applied. For early pulses precision at the TMS location was increased compared to sham, as well as compared to items probed contralateral to the TMS location. **(B)** When pulses were delivered early during the retention interval, participants were more likely to forget memory items, irrespective of their retinotopic location relative to the pulses. Parameter estimates were obtained by finding the best-fitting circular Gaussian (centered on 0° error of report, based on the mixed model analysis) for the frequency distribution of each condition, using a bin width of 12° (mean R^2 = 0.894 ± 0.029). Bin size was chosen to maximize the mean R^2 values across the different experimental conditions. Data were collapsed across hemispheres (left and right stimulation) before fitting in order to achieve a large enough number of trials per condition to obtain reliable fits. Error bars depict ± 1 SEM.

Similar to the absolute response errors, we wanted to directly assess memory variability at the TMS location while ignoring potential contributions of the TMS pulses across the different visual field locations. We fit the mixture-model to the combined responses at all three non-TMS locations, and contrasted this variability with that observed at the TMS ('same') location (circular *SD* from non-TMS locations minus circular *SD* at TMS location – Figure 5A). When pulses were delivered directly at stimulus offset, TMS resulted in increased memory precision (two-sided $t_{(7)}$ = 2.588; p = 0.036). Next we compared the two conditions during which neural effects were not expected: Memory variability during

sham trials was subtracted from memory variability on TMS trials during which locations non-overlapping with the pulses were probed. As expected, memory precision during sham trials did not differ from precision during TMS trials with items probed at non-TMS locations (both two-sided $p > 0.509$; Figure 5B).

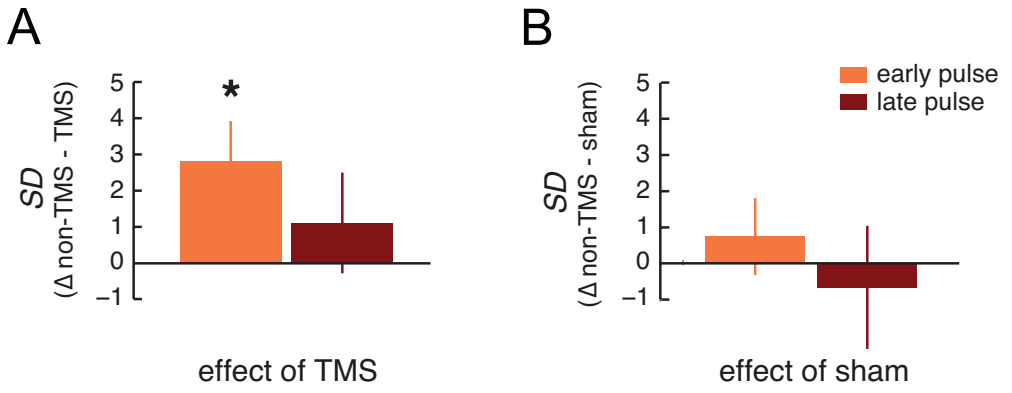


Figure 5. Difference metric of the mixed-model SD. **(A)** Difference between memory variability at the TMS-targeted location versus all the other locations probed during TMS trials. Positive difference scores indicate more precise memories at the TMS location compared to non-TMS locations. Mnemonic resolution was higher when the TMS pulses overlapped with the location of the memorized orientation and the TMS pulses were applied early during the retention interval. **(B)** Two conditions without an expected neural substrate were compared by subtracting memory variability on sham trials from memory variability on TMS trials during which stimuli were probed at locations not targeted by TMS. No differences in precision between sham and non-TMS locations were found, as indicated by difference scores around zero. Error bars depict ± 1 SEM.

Finally, the sham data were subjected to the same mixture-model analyses as the real TMS data. As already noted, pulse timing was considered the sole condition with potential behavioral consequences, meaning that too little trials were available to reliably perform this analysis split across all possible factors (absolute visual field location, pulse timing, stimulated hemisphere). Instead, the mixture model analysis was performed for each factor separately, collapsing across all others. We found no differences between the four absolute visual field locations for either memory precision ($F_{(3,21)} = 0.976$; $p = 0.423$) nor the probability of uniform responses ($F_{(3,21)} = 1.272$; $p = 0.310$). Timing of the sham pulses

did not affect memory variability (paired-samples $t_{(7)}$ = 0.822; p = 0.438) nor the probability of uniform responses (paired-samples $t_{(7)}$ = 1.262; p = 0.247). Finally, the hemisphere over which the sham coil was placed (left or right) also did not impact memory variability (paired-samples $t_{(7)}$ = 0.927; p = 0.385) nor the probability of uniform responses (paired-samples $t_{(7)}$ = 1.745; p = 0.124).

Memory performance across the visual field. A strength of our experimental setup was that the TMS coil was fixed over the skull, which, in combination with fMRI guided neuronavigation, ensured that the targeted brain site remained stable relative to the four patches of retinotopic cortex excited by our stimuli. However, this setup had the obvious side effect that the stimuli were anchored onto four static visual field locations across all experimental trials. If perception and memory at these static visual field locations were anisotropic, this would complicate interpretation of our results. One way by which we circumvented always targeting the same visual field location with TMS was by stimulating both the left and right hemispheres in a counterbalanced fashion. However, due to the folding of cortex TMS can only ever be administered to the dorsal part of visual cortex corresponding with the lower visual field, prohibiting a fully counterbalanced design. From the literature on basic human vision it is known that people are generally better at performing a visual task on stimuli in the lower, compared to the upper visual field. Luckily, such anisotropies are generally found for stimuli presented along the cardinal meridians, and less prevalent (or even absent) for stimuli presented at the obliques (Abrams, Nizam, Carrasco, 2012; Rovamo, Virsu, Laurinen, & Hyvärinen, 1982). Thus, while our stimuli were unlikely affected any residual anisotropy at the four stimulus locations would impact our results by biasing performance in favor of the lower half of the visual field, incidentally, the same half of the visual field targeted with TMS.

To preclude effects due to visual field anisotropies, one crucial check was to investigate how memory on sham trials varied across the stimulus locations in the upper left, upper right, lower left, and lower right parts of the visual field. These results were also discussed above: During sham neither the absolute error ($F_{(3,21)}$ = 0.459; p = 0.714), the precision of

memory ($F_{(3,21)}$ = 0.976; p = 0.423), nor the probability of uniform responses ($F_{(3,21)}$ = 1.272; p = 0.310) varied as a function of visual field location probed. While this suggests that visual field anisotropies cannot explain our main findings, null effects of sham are not the strongest possible argument. Directly comparing performance at the 'same' and 'contralateral' locations during real TMS would not very informative: A lack of differences between the two locations could be interpreted as a visual field effect (i.e. generally better performance in the lower visual field), or as a spread of the TMS pulses to the other hemisphere (i.e. TMS affecting both ipsi– and contralateral lower visual field locations).

Instead, we directly probed whether real TMS and sham might interact, as such an interaction would provide the most convincing evidence against the idea that upper versus lower visual field anisotropies (rather than TMS effects) might be driving performance differences at the four locations. In order to perform this direct comparison, we analyzed the absolute performance during sham in a manner identical to that of real TMS: performance was determined at each of the four visual field locations relative to the location 'targeted' by sham (for example, a sham coil over the right hemisphere 'targeted' the lower left visual field). Data for both sham and real TMS were collapsed across stimulated hemisphere, resulting in 160 and 40 observations per condition on average for real and sham TMS respectively. An ANOVA on the 2 TMS conditions (real and sham), 2 pulse timings (early and late), and 4 locations ('same', 'contra', 'ipsi', and 'diagonal') revealed an interaction between pulse timing and location ($F_{(3,21)}$ = 3.752; p = 0.027). Note that neither sham and location ($F_{(3,21)}$ = 0.204; p = 0.893) nor sham and timing ($F_{(1,7)}$ = 2.007; p = 0.2) interacted significantly. Post-hoc comparisons showed a marginally significant interaction ipsilateral to the pulse ($F_{(1,7)}$ = 5.077; p = 0.059), indicating that at this location real TMS pulses applied early decreased performance relative to sham, while real TMS pulses applied late increased performance relative to sham. While visual inspection of Figure 6 would suggest something similar happening at the 'same' location targeted directly with TMS, this interaction did not reach significance ($F_{(1,7)}$ = 0.934; p = 0.366). Finally, the apparent performance deficit for real compared to sham TMS for early pulses failed to reach significance ($F_{(1,7)}$ = 2.152; p = 0.186).

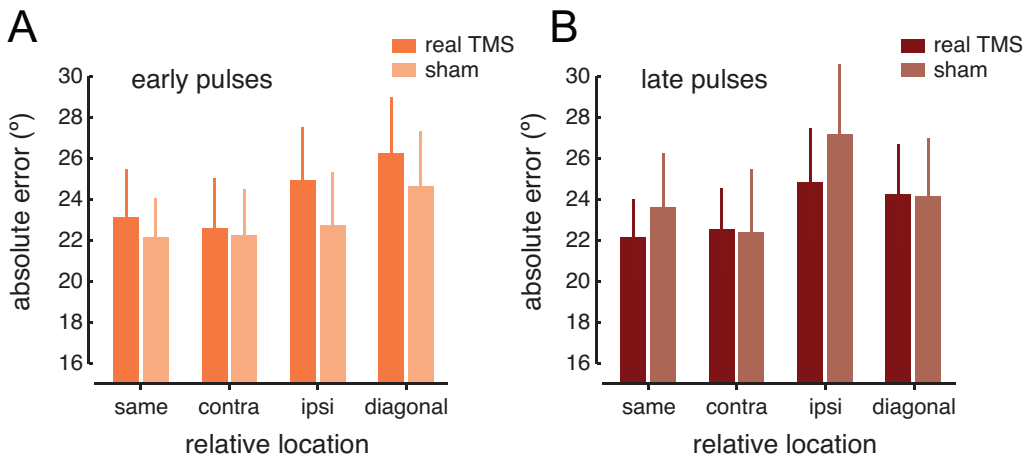


Figure 6. Absolute working memory performance during real and sham TMS. **(A)** Performance when real or sham pulses were applied early during retention, directly at stimulus offset. Data trends indicated worse performance when real pulses were applied early, compared to sham. **(B)** Performance for pulses (real or sham) applied midway through the retention interval. At the location ipsilateral to that targeted by the TMS pulses, data trends imply that TMS and sham have different effects on performance, with a negative and positive influence for early and late pulses respectively. Error bars depict ± 1 SEM.

Direct comparison of real and sham TMS did not yield entirely conclusive results. One obvious problem is that sham data were noisy, due to the smaller number of trials, making comparisons less robust. Nevertheless, the data suggest that real TMS and sham differentially affect performance, especially at visual field locations ipsilateral to the location targeted by the pulses. In fact, when we only considered visual field locations where real TMS would reasonably be expected to yield an effect (i.e. overlapping with and ipsilateral to the targeted location) the interaction between TMS condition and timing did reach significance ($F_{(1,7)} = 5.975$; $p = 0.044$).

Discussion

While participants were keeping four orientations in memory, 10Hz triple-pulse TMS was applied over early visual cortex retinotopically corresponding to the location of one of the remembered items. Memory for orientation differed between the four locations at which stimuli had been presented, with better performance at the location targeted by TMS compared to the location diagonal to TMS. Additionally, memory was impaired for early (directly at stimulus offset) compared to late pulses (midway through retention), irrespective of stimulus location. Decomposing replication errors into two summary statistics demonstrated the relative contributions of changes in memory variability on the one hand, and the probability of guessing on the other: Spatially specific improvements proximal to the pulse were attributed to reduced response variability, implying that memory precision can be improved locally by means of TMS. Global impairments imposed by early compared to late pulses were due to an increased likelihood of random responses, implying retinotopically aspecific disturbances due to TMS at the tail end of encoding. None of these TMS findings were present during sham, signifying the neural basis underlying the effects reported here. Moreover, while contributions from anisotropies across the different stimulus positions in the visual field cannot be excluded, these were unlikely driving all of the observed effects.

Our task encouraged binding between memorized features and their positions in the visual field by making memory retrieval contingent upon the spatial location of stimuli. The spatial extent of sensory recruitment has been a matter of some debate, found in some cases (Pratte & Tong, 2014) but not others (Ester et al., 2009). Our findings of locally improved memory performance and precision provide support for retinotopically specific sensory recruitment during visual memory. This was corroborated by contrasting performance and precision at the TMS-targeted location with all other locations in our experiment, showing enhanced outcomes at the TMS location. By contrast, a similar comparison between sham and non-TMS sites (two conditions without neural substrates) did not reveal any differences.

A higher likelihood of random responses for early compared to late pulses implied that different processes occurred at different phases of the retention interval. This finding could not be explained by simple modulations of attentional cuing or distractibility due to the timing of the three auditory clicks emitted by the TMS coil – using similar timings and sounds revealed no costs for early compared to late sham-pulses. So why might early pulses lead to an increase in random responses, and why is this happening across the visual field? First of all, uniform responses can be a consequence of forgetting, but also of factors such as lapses of attention during stimulus presentation, or encoding failures. Since early pulses were presented at the tail end of encoding (Chun & Potter, 1995; Jolicoeur & DellAcqua, 1998; Vogel, Woodman, & Luck, 2006), we believe it's likely that TMS increased random guesses by prohibiting adequate encoding of the four stimuli. Similarly, previous work showed that TMS disrupted visual memory when applied over visual cortex during the early stages of retention (van de Ven et al., 2012). However, this disruption was observed in a retinotopically specific manner. Why might we find that early pulses affected performance all across the visual field?

While stimulus orientations were chosen randomly and independently, accidental yet strong ensemble effects probably occurred on a portion of trials by virtue of our design. Restricting the four orientations to the same locations on every trial, participants could have adapted a strategy relying on the constellation of the four orientations (as radial, concentric, isotropic, etc.) rather than storing features in a truly independent manner. Perceptual grouping of elements allows more of them to be stored in memory (Luck & Vogel, 1997), and taking perceptual grouping and higher-order structures between items into account is a requirement when attempting to explain memory performance (Brady, Konkle, & Alvarez, 2009; Brady & Tenenbaum, 2013). Thus, early pulses applied while participants were extracting global shape-like representations could disrupt encoding of the whole 'object' through the disturbance of localized 'object features' (i.e. orientations), resulting in increased guesses for all orientations during retrieval. The 'object-based' working memory strategy suggested here might be achieved by convergent feed-forward and feedback processes at multiple stages of the visual hierarchy (Lamme & Roelfsema,

2000; Pooresmaeili & Roelfsema, 2014) – an intriguing possibility that could be tested empirically in the future.

People generally conceive of TMS as a way to examine the 'necessity' of a given brain area in a given task. By inflicting a 'virtual lesion', hopes are to find performance decrements that in turn permit the conclusion that a targeted area is critically involved in the task at hand. In line with this general idea, our initial hypothesis was that early visual cortex TMS during working memory retention would disrupt performance, and cause noisier stimulus representations. Our findings attested to the contrary, with TMS improving rather than impairing memory precision at the targeted location. Prior studies applying TMS over visual cortex during working memory maintenance have yielded a mixed bag of results, unveiling both performance *decrements* (Cattaneo, Vecchi, Pascual-Leone, & Silvanto, 2009; van de Ven et al., 2012; Silvanto & Soto, 2012), as well as *improvements* (Cattaneo et al., 2009; Soto, Llewelyn, & Silvanto, 2012). Variable outcomes like this can be due to many factors, amongst which the details of the task. Unique to our design was the simultaneous presentation of stimuli in all four quadrants of the visual field. This raises the possibility that, while participants maintained multiple representations at distributed locations, TMS acted to draw resources to the targeted retinotopic location in a way equivalent to the drawing of attentional resources to a particular location.

One straightforward way by which TMS might mimic spatial attention could be by enhancing neural activity at a retinotopically specific location. It has been shown that in the absence of a visual stimulus, spontaneous firing rates in V2 ad V4 were elevated when attention was directed at a location that fell within a cell's receptive field (Luck, Chelazzi, Hilyard, Desimone, 1997). Thus, TMS during memory could instigate processes mimicking the effects achieved by the brain when employing top-down spatial attention (Luck, et al., 1997), potentially by means of a spatial gating mechanism that favors the boosted location during subsequent processing stages (Chen & Seideman, 2012). This would suggest that TMS acts as a bottom-up implementation of an otherwise top-down biasing signal. In this hypothesis, the effects of TMS reach beyond simply mimicking

exogenous attention, as the processing benefits enjoyed by mere bottom-up transients are short-lived (Nakayama & Mackeben, 1989) and therefore less likely to have acted within the time span of our paradigm.

At a more computational level, TMS could have mimicked the deployment of spatial attention by attenuating regulatory processes that mediate competition between orientation representations at various locations in the visual field. Specifically, unanticipated inhibitory interactions might have ensued between the four simultaneously remembered orientations at a neural level. Lateral connections between neurons in posterior visual areas with relatively large aggregate receptive field sizes (Freeman & Simoncelli, 2011), monosynaptic trans-colossal connections between the primary visual hemispheres (Berlucchi & Rizzolatti, 1968), or top-down influences from for example prefrontal cortex (Bays, 2015; Edin et al., 2009) are all routes through which interactions between multiple and spatially distributed stimulus representations might arise. It has been shown that, at a local level, TMS can release orientation representations from inhibitory interactions during a tilt illusion paradigm (Ling, Pearson, & Blake, 2009). Likewise, in our study TMS could have similarly acted to depress horizontal connections between the four stimuli on a more global level, attenuating interference of the other stimuli at the targeted location.

This idea is corroborated by recent work suggesting a population-coding model of working memory (Bays, 2015; Sreenivasan, Curtis, & D'Esposito, 2014; Stokes, 2015), which assumes that stimulus features (like orientation) are stored by probabilistic spiking activity in tuned populations neurons. Critically, this model includes a broad normalization component by assuming that the sum of the firing rates remains constant across changes in set size or changes in attentional prioritization. Normalization is a computation that describes neuronal responses as consisting of a neuron's own input or 'drive', divided by the summed activity of other neurons in a normalization pool (Carandini & Heeger, 2011; Heeger, 1992). The assumption that normalization occurs between all items held in memory could allow the population-coding model to explain our

findings by a multiplicative increase of the input drive of neurons at the site targeted with TMS, biasing the overall population activity.

Finally, we would like to suggest a manner by which TMS might have enhanced processing of a memorized orientation locally, without pertaining to tentative interactions between simultaneously remembered stimuli. The first basic premise is that TMS acts via a wholesale multiplication of neural responses. The second premise is that while remembering an orientation, the memory trace of that orientation in a population of orientation selective neurons is weak, with firing rates only slightly elevated above baseline. In terms of intensity response, the remembered orientation has a response that is slightly larger than that of not-remembered orientations. However, because of the nonlinearity of intensive response profiles, a wholesale multiplication by any factor due to TMS would result in higher signal-to-noise for the remembered orientation. Conceivably, such signal-to-noise benefits could be induced with TMS a directly perceived orientation as well, as long as contrast remains lower that the inflection point of the contrast response function, and intensive responses undergo expansive nonlinearity.

Surely this list of possible means by which TMS could improve memory is not exhaustive, but it is post-hoc, and these hypotheses should be tested empirically to ascertain their true value. To that, we'd like to add some additional considerations regarding the work presented here. TMS effects are usually small, and people's responses to TMS highly variable, which is why large sample sizes are generally desirable. Instead, here we opted for testing a smaller number of participants in depth over very many trials. While psychophysically rigorous, this approach may not be optimally suited for a TMS context. Thus, the strength of our design, allowing us to investigate the mechanisms supporting memory maintenance in visual cortex, is simultaneously its weakness. One consequence was that our data did not provide conclusive evidence regarding the effects of TMS compared to sham in the context of possible visual field anisotropies. Another that we could not be certain whether we observed a true TMS-induced improvement, or a mixture of both improved and impaired performance more proximal and distal to the pulses

respectively.

Another weakness of combining small effect sizes inherent to TMS with comprehensive psychophysical testing is that noisy responses from only a few participants precluded reliable fitting using Maximum Likelihood Estimation. Instead we binned the data in bins of 12° before deriving parameter estimates. While rather coarse, this bin size was chosen to maximize R^2 values, and results were comparable with smaller bins (i.e. 8° or 10°). Error fitting performed in this way is purely empirical, unlike model fitting, and could be considered the better choice here. However, it should be noted that when applying a maximum likelihood approach the directionality of effects was preserved, but neither precision differences across the visual field ($F_{(3,21)}$= 1.837; p = 0.171), nor the pulse timing effects on guessing responses ($F_{(1,7)}$ = 2.854; p = 0.135) remained significant.

Despite these cautionary notes, our results provide several consistent and intriguing findings. Moreover, a simple mixture-model approach in combination with TMS offers novel insights into potential sources underlying memory errors when that memory is perturbed. Specifically, it permits a glimpse into the kinds of processes supported by early visual areas during memory, while affirming the involvement of these areas during the retention of visual items. We were able to uncover a double dissociation, showing local changes in memory precision irrespective of pulse timing, but global changes in guess rates contingent on pulse timing, only at the tail end of encoding. As such, our work adds to the existing literature demonstrating retinotopically specific sensory recruitment (Pratte & Tong, 2014) and studies utilizing TMS over early visual cortex to investigate the mechanisms of visual memory (van de Ven et al., 2012).

References

Abrams, J., Nizam, A., & Carrasco, M. (2012). Isoeccentric locations are not equivalent: the extent of the vertical meridian asymmetry. *Vision Research*, *52*(1), 70–78.

Albers, A.M., Kok, P., Toni, I., Dijkerman, H.C., & de Lange, F.P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*(15), 1427–1431.

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

Bays, P.M. (2014). Noise in Neural Populations Accounts for Errors in Working Memory. *Journal of Neuroscience*, *34*(10), 3632–3645.

Bays, P.M. (2015). Spikes not slots: noise in neural populations limits working memory. *Trends in Cognitive Sciences*, *19*(8), 431–438.

Bays, P.M., Catalao, R.F.G., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10):7, 1–11.

Bays, P.M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*(5890), 851–854.

Berlucchi, G., & Rizzolatti, G. (1968). Binocularly driven neurons in visual cortex of split-chiasm cats, Science, 159(3812), 308–310.

Bosch, S.E., Jehee, J.F.M., Fernández, G., & Doeller, C.F. (2014). Reinstatement of associative memories in early visual cortex is signaled by the hippocampus. *The Journal of Neuroscience*, *34*(22), 7493–7500.

Boynton, G.M., Engel, S.A., Glover, G.H., & Heeger, D.J. (1996). Linear systems analysis of Functional Magnetic Resonance Imaging in human V1. *The Journal of Neuroscience, 16*(13), 4207–4221.

Brady, T.F., Konkle, T., & Alvarez, G.A. (2009). Compression in visual working memory: using statistical regularities to form more efficient memory representations. *Journal of Experimental Psychology General*, *138*(4), 487–502.

Brady, T.F., & Tenenbaum, J.B. (2013). A probabilistic model of visual working memory: Incorporating higher order regularities into working memory capacity estimates. *Psychological Review*, *120*(1), 85–109.

Buschman, T.J., Siegel, M., Roy, J.E., & Miller, E.K. (2011). Neural substrates of cognitive capacity limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(27), 11252–11255.

Carandini, M., & Heeger, D.J. (2011). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 1–12.

Cattaneo, Z., Vecchi, T., Pascual-Leone, A., & Silvanto, J. (2009). Contrasting early visual cortical activation states causally involved in visual imagery and short-term memory. *The European Journal of Neuroscience*, *30*(7), 1393–1400.

Chen, Y., & Seidemann, E. (2012). Attentional modulations related to spatial gating but not to allocation of limited resources in primate V1. *Neuron*, *74*(3), 557–566.

Christophel, T.B., Hebart, M.N., & Haynes, J.-D. (2012). Decoding the contents of visual short-term memory from human visual and parietal cortex. *The Journal of Neuroscience*, *32*(38), 12983–12989.

Chun, M.M., & Potter, M.C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology*, *21*(1), 109–127.

Courtney, S.M., Ungerleider, L.G., Keil, K., & Haxby, J.V. (1996). Object and spatial visual working memory activate separate neural systems in human cortex. *Cerebral Cortex*, *6*(1), 39–49.

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society of London Series B, Biological Sciences*, *362*(1481), 761–772.

Edin, F., Klingberg, T., Johansson, P., McNab, F., Tegner, J., Compte, A. (2009). Mechanisms for top-down control of working memory capacity. *Proceedings of the National Academy of Sciences of the United States of America, 106*(16), 6802–6807.

Emrich, S.M., Riggall, A.C., LaRocque, J.J., & Postle, B.R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. *The Journal of Neuroscience*, *33*(15), 6516–6523.

Ester, E.F., Serences, J.T., & Awh, E. (2009). Spatially global representations in human primary visual cortex during working memory maintenance. *The Journal of Neuroscience*, *29*(48), 15258–15265.

Ester, E.F., Sprague, T.C., & Serences, J.T. (2015). Parietal and Frontal Cortex Encode Stimulus- Specific Mnemonic Representations during Visual Working Memory. *Neuron*, 1–30.

Fougnie, D., Suchow, J.W., & Alvarez, G.A. (2012). Variability in the quality of visual working memory. *Nature Communications*, *3*, 1229.

Freeman, J., & Simoncelli, E.P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, *14*(9), 1195–1201.

Friston, K.J., Holmes, A.P., Poline, J.B., Grasby, P.J., Williams, S.C., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *NeuroImage*, *2*(1), 45–53.

Harris, J.A., Miniussi, C., Harris, I.M., & Diamond, M.E. (2002). Transient storage of a tactile memory trace in primary somatosensory cortex. *The Journal of Neuroscience*, *22*(19), 8720–8725.

Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Heeger, D.J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*, 181–197.

Hollingworth, A. (2006). Scene and Position Specificity in Visual Memory for Objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(1), 58–69.

Hollingworth, A. (2007). Object-position binding in visual memory for natural scenes and object arrays. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(1), 31–47.

Horikawa, T., Tamaki, M., Miyawaki, Y., & Kamitani, Y. (2013). Neural decoding of visual imagery during sleep. *Science*, *340*(6132), 639–642.

Jolicoeur, P., & Dell'Acqua, R. (1998). The demonstration of short-term consolidation. *Cognitive Psychology*,

36(2), 138–202.

Jonides, J., Lacey, S. C., & Nee, D. E. (2005). Processes of working memory in mind and brain. *Current Directions in Psychological Science, 14*(1), 2–5.

Kosslyn, S.M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J.P., Thompson, W.L., et al. (1999). The role of area 17 in visual imagery: convergent evidence from PET and rTMS. *Science*, *284*(5411), 167–170.

Lamme, V.A., & Roelfsema, P.R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *TRENDS in Neurosciences*, *23*(11), 571–579.

Lee, S.-H., Kravitz, D.J., & Baker, C.I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073.

Lee, S.-H., Kravitz, D.J., & Baker, C.I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature*, *16*(8), 997–999.

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., & Vogel, E.K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*(6657), 279–281.

Meyer, K., Kaplan, J.T., Essex, R., Webber, C., Damasio, H., & Damasio, A. (2010). Predicting visual stimuli on the basis of activity in auditory cortices. *Nature Neuroscience*, *13*(6), 667–668.

Nakayama, K., & Mackeben, M. (1989). Sustained and transient components of focal visual attention. *Vision Research, 29*(11), 1631–1647.

O'Craven, K., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023.

Pasternak, T., & Greenlee, M.W. (2005). Working memory in primate sensory systems. *Nature Reviews Neuroscience, 6*, 97–107.

Pasternak, T., & Zaksas, D. (2003). Stimulus specificity and temporal dynamics of working memory for visual motion. *Journal of Neurophysiology*, *90*(4), 2757–2762.

Pooresmaeili, A., & Roelfsema, P.R. (2014). A Growth-Cone Model for the Spread of Object-Based Attentionduring Contour Grouping. *Current Biology*, *24*(24), 2869–2877.

Pratte, M.S., & Tong, F. (2014). Spatial specificity of working memory representations in the early visual cortex. *Journal of Vision*, *14*(3):22 1–12.

Riggall, A.C., & Postle, B.R. (2012). The relationship between working memory storage and elevated activity as measured with functional magnetic resonance imaging. *The Journal of Neuroscience, 32*(38), 12990–12998.

Rovamo, J., Virsu, V., Laurinen, P., & Hyvärinen, L. (1982). Resolution of gratings oriented along and across meridians in peripheral vision. *Investigative Ophthalmology & Visual Science*, *23*(5), 666–670.

Serences, J.T., & Boynton, G.M. (2007). Feature-based attentional modulations in the absence of direct visual stimulation. *Neuron*, *55*(2), 301–312.

Serences, J.T., Ester, E.F., Vogel, E.K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214.

Silvanto, J., & Cattaneo, Z. (2010). Transcranial magnetic stimulation reveals the content of visual short-term memory in the visual cortex. *NeuroImage*, *50*(4), 1683–1689.

Silvanto, J., & Pascual-Leone, A. (2008). State-Dependency of Transcranial Magnetic Stimulation. *Brain Topography*, *21*(1), 1–10.

Silvanto, J., & Soto, D. (2012). Causal evidence for subliminal percept-to-memory interference in early visual cortex. *NeuroImage*, *59*(1), 840–845.

Sneve, M. H., Sreenivasan, K. K., Alnæs, D., Endestad, T., & Magnussen, S. (2015). Short-term retention of visual information: Evidence in support of feature-based attention as an underlying mechanism. *Neuropsychologia*, *66*(C), 1–9.

Soto, D., Llewelyn, D., & Silvanto, J. (2012). Distinct causal mechanisms of attentional guidance by working memory and repetition priming in early visual cortex. *The Journal of Neuroscience*, *32*(10), 3447–3452.

Sprague, T.C., Ester, E.F., & Serences, J.T. (2014). Reconstructions of Informationin Visual Spatial Working Memory Degrade with Memory Load. *Current Biology*, *24*(18), 2174–2180.

Sreenivasan, K.K., Curtis, C.E., D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences, 18*(2), 82–89.

Stokes, M.G. (2015). 'Activity-silent' working memory in prefrontal cortex: a dynamic coding framework. *Trends in Cognitive Sciences, 19*(7), 394–405.

Todd, J.J., & Marois, R. (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, *428*(6984), 751–754.

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579.

van den Berg, R., Shin, H., Chou, W.-C., George, R., & Ma, W.J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(22), 8780–8785.

van de Ven, V., Jacobs, C., & Sack, A.T. (2012). Topographic Contribution of Early Visual Cortex to Short-Term Memory Consolidation: A Transcranial Magnetic Stimulation Study. *The Journal of Neuroscience*, *32*(1), 4–11.

Vogel, E.K., Woodman, G.F., & Luck, S.J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436–1451.

Woodman, G.F., Vogel, E.K., & Luck, S.J. (2012). Flexibility in Visual Working Memory: Accurate Change Detection in the Face of Irrelevant Variations in Position. *Visual Cognition*, *20*(1), 1–28.

Xu, Y., & Chun, M.M. (2006). Dissociable neural mechanisms supporting visual short-term memory for objects. *Nature*, *440*(7080), 91–95.

Zaksas, D., Bisley, J. W., & Pasternak, T. (2001). Motion information is spatially localized in a visual working-memory task. *Journal of Neurophysiology*, *86*(2), 912–921.

Zhang, W., & Luck, S. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, *453*(7192), 233–235.

Chapter 6

Intensive tool-practice and skillfulness facilitate the extension of body representations in humans

Corresponding publication:

Rademaker R.L., Wu, D-A., Bloem I.M., Sack A.T. (2014).

Intensive tool-practice and skillfulness facilitate the extension of body representations in humans.

Neuropsychologia, 56, 196–203.

Abstract

The brain's representation of the body can be extended to include objects that are not originally part of the body. Various studies have found both extremely rapid extensions that occur as soon as an object is held, as well as extremely slow extensions that require weeks of training. Due to species and methodological differences, it is unclear whether the studies were probing different representations, or revealing multiple aspects of the same representation. Here, we present evidence that objects (cotton balls) held by a tool (chopsticks) are rapidly integrated into the body representation, as indexed by fading of the cotton balls (or 'second-order extensions') from a positive afterimage. Skillfulness with chopsticks was predictive of more rapid integration of the second-order objects held by this tool. We also found that extensive training over a period of weeks augmented the level of integration. Together, our findings demonstrate integration of second-order objects held by tools, and reveal that the body representation probed by positive afterimages is subject to both rapid and slow processes of adaptive change.

Introduction

Imagine a skillful tennis player immersed in a heated match requiring his upmost capacity and focus. For an external observer, the tennis player is typically considered an independent actor and cause of the events he initiates within his surrounding environment. But in the tennis-player's mental experience, his body, the racket and even the ball can be felt as part of his sensory and intentional self. As the ball approaches, his thoughts are less likely to be on the desired trajectory of his arm, than on the trajectory of the racket head. When the racket makes contact with the ball, the feeling of impact is perceived not at the tactile sensors in his hand, but in the racket head itself. At high levels of skill and concentration, even the racket may become secondary in his experience, all thoughts becoming based on the ball and its desired trajectory. This ability for conscious awareness to be focused on the ball requires that the intermediate effectors (muscles, joints, racket) be integrated into a subconscious, automatically processed model. This model must be capable of tracking the current states of the effectors, and of back-calculating conscious goals into basic motor commands.

The original positing of a model representing the body came from studies of neurological patients by Head & Holmes (1911). Based on observed deficits in postural awareness and tactile localization, they proposed that the normally functioning brain has two types of bodily representations. First there is the *body image*, a conscious representation that is the subject of our thoughts and perceptual judgments. Secondly, there is the *body schema*, an unconscious framework that automatically integrates posture, proprioceptive input and action goals into a common spatial frame.

The *body image* is believed to be a multisensory representation of the body that integrates stored knowledge, and by subserving mainly perceptual purposes it is subject to bodily illusions (Kammers, Kootker, Hogendoorn, & Dijkerman, 2010). For example, vibrations applied to a tendon causing the sensation of that tendon stretching will result in the perceptual experience of the corresponding limb being moved (Goodwin, McCloskey, &

Matthews, 1972). Another manipulation of the body image is demonstrated by the 'rubber hand illusion'. Here, sensory conflict is induced by simultaneous stroking of the own (unseen) hand and a visible rubber hand, resulting in an experience of tactile sensations occurring at the rubber hand (Botvinick & Cohen, 1998).

The *body schema*, on the other hand is described as an unconscious representation that subserves action rather than perception. Head & Holmes proposed that this schema does not exclusively code for the physical body, but is capable of extending to objects that are needed to support skilled actions or smooth movement through the environment. Thus, the body schema would need to include tools, or even a large feather in one's hat, in order to support one's actions or avoid collisions. Though generally believed to be highly robust, also this motoric body representation is not entirely immune to bodily illusions. For example, after inducing the rubber hand illusion, the grip aperture of a real hand was found to mimic that of a rubber hand (Kammers, et al., 2010).

Since the early work of Head & Holmes, people largely agree on the existence of multiple body representations, though their exact number and definition is still a matter of debate (Cardinali, et al., 2009; Cardinali, et al., 2012; de Vignemont, 2010; Kammers, et al., 2010).

Evidence that tools become integrated into these body representations has come via various experimental routes. Changes in the *body schema* are most directly observed by monitoring the kinematics of action execution. In a study by Cardinali et al (2009), participants who used a mechanical grabber subsequently changed the kinematics of their empty-handed movements, pointing and grasping as if their arms had lengthened. Simple motor learning was an unlikely account for these changes, as the kinematics of tool-use itself did not change throughout the period in which the mechanical grabber was used. Given that tool-use induced changes in empty-handed actions, the results suggested that a change had occurred in a generalized model of action generation. These findings therefore imply a highly plastic representation of the body schema, similar to what had been suggested by Head & Holmes almost a century prior.

The other major class of tool-use experiments uses measures of multimodal integration to investigate body representations (Maravita & Iriki, 2004). Certain sensory processes are selective for stimuli originating from within "peripersonal space", which corresponds to the reachable or "actionable" space immediately surrounding the body. Bodily representations both define the extent of this space, and also form a basis for the spatial mapping of sensory stimuli within it. Thus, monitoring changes in the extent and organization of this sensory space allows one to infer changes in body representations.

A lot of what is known about body representations in peripersonal space comes from neurophysiological studies in primates. Fronto-parietal networks have been identified that continuously update spatial representations of body shape and posture. These networks integrate multimodal sensory information (primarily proprioceptive, somatosensory and visual information) such that it is functionally relevant to specific actions, and serves the ability to localize the body in space (Colby, 1998; Maravita, Spence, & Driver, 2003; Rizzolatti, Fadiga, Fogassi, & Gallese, 1997). Notably, there are neurons in ventral-premotor cortex that have both somatosensory and visual receptive fields, coding for the space surrounding the same body part. These bimodal neurons integrate information such that even if a body part (for example a hand) is moved through space, the visual receptive field remains anchored to the body part it belongs to (Graziano, Yap, & Gross, 1994).

Intriguingly, these fronto-parietal networks can represent external objects in a similar fashion. After weeks of practice with a simple tool, bimodal neurons in intraparietal cortex of macaques were found to expand their visual receptive fields to include the space surrounding the tool, while the monkey was engaged in deliberate tool interaction (Iriki, Tanaka, & Iwamura, 1996). This finding suggests that peripersonal space can be expanded via the use of a tool (but see also Holmes, 2012). Similarly, a study investigating structural brain changes in macaques exposed to tool-use training for the first time, showed an increase in grey matter volume in fronto-parietal areas including intraparietal cortex (Quallo, et al., 2009). In a study of a human patient with right-hemisphere lesions, tool-use altered the domain in which visual neglect was experienced. Whereas the patient's

visual neglect was typically restricted to judgments regarding stimuli in peripersonal space, the neglect spread to more distant areas if the task was performed using a long pointing tool, again suggestive of the expansion of peripersonal space (Berti & Frassinetti, 2000). Increased multisensory weights assigned to the processing of visual stimuli around the *functional* part of a tool are likely responsible for the remapping of peripersonal space to include this new region of space after tool-use (Holmes, 2012). Note that none of these studies probed motor output as a dependent measure, so it is unclear whether these body representations subserve action planning as a body schema, or if they subserve only perceptual processing.

The present study utilizes another method of probing bodily representations, which has recently been extended to investigate tool use. The paradigm involves a cross-modal effect whereby proprioceptive inputs profoundly disrupt visual representations of the body (Bross, 2000; Davies, 1973a; Gregory, Wallace, & Campbell, 1959). In these experiments, participants in a completely darkened room are exposed to a brief flash of light, which creates a crisp, long-lasting afterimage of the entire field of view. When the afterimage includes a body part, such as the participant's arm, moving the arm from its imaged position causes the afterimage of the arm to 'fade' or 'crumble' while the rest of the afterimage scene remains intact. The mismatch between proprioceptive and visual representations of the same body part leads to a Gestalt-like disruption of the visual percept. Versions of this experiment done with mirrors confirm that this fading effect occurs in accordance with proprioceptive and visual representations organized on the basis of one's own body (Ritchie & Carlson, 2010).

Such afterimage-based experiments have also demonstrated the rapid modulation of body representations to include held objects. Carlson, Alvarez, Wu & Verstraten (2010) showed that objects grasped by the observer (referred to as *'first-order' objects*) faded upon being dropped. Similarly, when the observer removed a first-order object from the area of peripersonal space being viewed in the afterimage, the object would also fade. This

indicates that somatosensory and proprioceptive information is integrated with visual information in much the same way for both held objects and body parts.

Afterimage studies do not investigate motor output, and thus the body representations that were probed may or may not function as body schema. The representations seem more clearly akin to the ones probed in the studies of peripersonal space. Both involve multisensory integration and measurements based on perceptual outcomes. Using the afterimage paradigm, we aim to address several related issues raised by the preceding studies. What kinds of external objects are assimilated? What factors govern whether or not an object is assimilated? How quickly does assimilation occur?

Although the monkey physiology studies found that tool integration developed after weeks of use (Iriki, et al., 1996), the human behavioral studies found tool integration as soon as the tools were grasped (Cardinali et al., 2009; Carlson et al., 2010). The behavioral findings closely match our daily functioning and the feeling that we can rapidly assimilate objects (like picking up a pen and beginning to write). There are many functional advantages to a body system capable of rapidly incorporating, as well as disincorporating, an object or tool. The ability to readily expand the physical body in a functional manner via tool-use enables us to do numerous things that would not otherwise be possible, such as removing hot coals from a fire or hitting a nail into a wall. Being able to readily disincorporate a tool after it is no longer used allows us to keep a coherent sense of the body's boundaries. Such short-lived changes in the brain's representation of the body might be most effectively established by flexibly updating representations of peripersonal space (Bruggeman, Kliman-Silver, Domini, & Song, 2013; Carlson, et al., 2010; Holmes, 2012). However, extending the body into space for functional purposes could also involve updating of the action oriented body schema (Cardinali, et al., 2009).

Such a flexible system could be beneficial for the incorporation of tools, but also for objects held by tools. Certainly we do many things involving second, or even higher order extensions. Extreme examples of this are operating construction vehicles, performing

robotic surgery, etc. But there are also many more low-tech examples, such as the use of chopsticks to manipulate food while eating a meal. Taking this marked degree of flexibility, and the goal-oriented nature of tool-use and body representations into account, one could readily anticipate representations for higher-order extensions.

However, the afterimage experiments revealed complex results regarding second-order extensions. When participants used a simple, table-supported, mechanical arm that could grip objects when squeezing a handle, objects held by the tool did not fade from the afterimage when participants released the arm's handle (Bruggeman, et al., 2013; Carlson, et al., 2010). This finding demonstrates that an observer's ability to consciously predict the consequences of an action (e.g. a hand movement leading to release of an object) is not sufficient to induce fading of an object from the afterimage.

But in light of the sensitivity that the body's representation has for afferent sensory input (Carlson, et al., 2010; Hogendoorn, Kammers, Carlson, & Verstraten, 2009) the apparent failure to integrate higher-order external objects seems counter to our daily experience. In fact, the study by Bruggeman et al. shows that also second-order objects can fade from the afterimage when released from a mechanical arm, as long as participants are able to freely wield the tool. Contrary to using a tool while it is fixed to a table, freely wielding the same tool offers rich somatosensory feedback, providing the information necessary to experience fading of a second-order object (Bruggeman, et al., 2013).

The critical factor, Bruggeman et al. suggest, is that target objects can be perceived directly, or indirectly via a tool, through 'dynamic touch'. Dynamic touch can be defined by the combined muscular effort and sensory consequences of manipulating an object (Gibson, 1966; Yamamoto & Kitazawa, 2001). Mechanoreceptors in the hand are able to detect mechanical forces (such as torque, and moment of inertia) that emerge when one manipulates a tool, and such signals (mainly the inertia tensor) can be used by the brain to quantify for example the length of a handheld object without requiring vision (Turvey, 1996). Dynamically manipulating an object involves both perception and action, allowing

the object to become incorporated into the action and somatosensory system (Bruggeman, et al., 2013).

The importance of action is also demonstrated by studies where the simple physical presence of a tool does not induce remapping of peripersonal space, which instead requires deliberate tool-action (Farnè, Iriki, & Làdavas, 2005; Iriki, et al., 1996; Wagman & Carello, 2001). Moreover, the ability to predict action outcomes is crucial for tool-use, and requires a tight link between motor predictions and feedback from the somatosensory system (Wolpert, Goodbody, & Husain, 1998; Wolpert & Flanagan, 2010). Thus, a flexible and quickly adaptive system consisting of a feedback loop between perception and action would be well suited to the demands of rapidly incorporating and disincorporating second-order objects and tools.

Given that (afterimage) studies in humans reveal a highly flexible and rapidly changing body representation, capable of assimilating even second-order objects, why do studies of macaque neurophysiology only reveal slowly changing representations? One explanation for this might be the existence of multiple representations that differ in learning speed, as proposed by Carlson, et al. (2010). Alternatively, differences in assimilation speed of objects could be due to the amount of sensorimotor feedback provided by a given tool (Bruggeman, et al., 2013). Finally, differing assimilation speeds found by various studies could stem from important differences between species. Whereas humans are dexterous tool-users, lower primates are not consistently known to engage in the spontaneous use of tools (Iriki, et al., 1996; Tomasello & Call, 1997). It's been suggested that, in macaques, training is needed to activate silent neurogenetic mechanisms (Ishibashi, et al., 2002b; Tomasello & Call, 1997), which represent a precursor of the tool-use abilities acquired by humans over the continued course of evolution (Ishibashi, et al., 2002a). This could account for the slower buildup needed to find integration of external tools into the body representation of monkeys.

At present, studies in human subjects have not tested for changes in tool integration across the long timescales present in the monkey studies. Conceivably, long-term practice could improve somatosensory perception of a tool, leading to better predictions of motor actions performed with that tool. Such slowly evolving improvements in dynamic touch may drive neural plasticity during tool-use, changing the extent to which second-order objects are assimilated. This would reveal whether plasticity in the human body representation underlying the fading effect contains a slow component.

In the present study, we use the afterimage paradigm to investigate tool integration during the use of chopsticks. Chopsticks provide rich somatosensory feedback during use, and are thus a good choice where dynamic touch is important. We first verify the existence of second-order integration by using chopsticks. This is not entirely a given, since chopsticks rely more on finger representation and kinesthesia, while the long mechanical grabbers used in previous research are likely to rely on information from more proximal parts of the hand and arm representations (Bruggeman, et al., 2013; Cardinali, et al., 2009; 2012; Carlson, et al., 2010).

Second, we test for both elements of rapid integration and build-up through training. In two successive experiments we assess whether skillfulness with a tool can modulate potential rapid integration, and whether extensive tool practice in humans can result in long-term integration of objects, similar to that found in monkeys. Throughout, we compare participants' dominant and non-dominant hands, which provides within-subject control. Also, this helps us answer a third possible question, since the degree of life experience with a tool needed to modulate the fading effect is an open question– handedness covers one extreme, since it is built up over one's entire lifetime.

Here, we demonstrate that second-order objects do fade fairly frequently, validating the idea that space representations around the body can be modified to include second-order objects. We find that such integration of second-order objects includes both a rapid component, as well as a longer-term component that can be built up over the course of

extensive training. These findings provide a link between the fast and highly flexible integration of objects found in many behavioral studies with humans, and the slower buildup of tool integration through training found in monkey studies.

Methods

Participants. Fifty-nine healthy volunteers were recruited from Maastricht University (35 female, 51 right-handed, mean age ~22 years). Data from eight participants were excluded due to missing audio-recordings (n=1), inability to hold chopsticks (n=1) and failing a pre-experimental screening procedure (see below, n=6). Eight volunteers, including two of the authors (RR and IB) also participated in experiment 2: a four-week training and post-training assessment. All participants had normal or corrected-to-normal vision, agreed to the use of voice-recordings and provided written informed consent. Participants were reimbursed by means of course credit. The study took place under the approval of the standing ethical committee of the Psychology and Neuroscience department at Maastricht University.

Stimuli and Procedure. Experiments were conducted in a completely darkened room. Participants were seated facing a table and wall covered by a black cloth (1.56 cd/m^2) extending ~60º of visual angle horizontally. Afterimages were created using a handheld Vivitar 285HV Zoom Thyristor flashgun directed at the ceiling. Participants' verbal reports were collected with an iRiver (iHP-120 multi-codec jukebox) device. Wooden chopsticks (24 cm in length, unfinished) were employed for all experimental operations: experimental trials, training and a skill test. The second-order objects were cotton balls dyed in black tea (88.3 cd/m^2), which were easily graspable with chopsticks and provided no auditory feedback when dropped.

Because our experiment depends on the ability of participants to experience object fading, we conducted a separate pre-experimental screening procedure assessing the ability of each participant to see changes in the positive afterimage in response to bodily movement

(Carlson et al., 2010). Participants were excluded from further participation if no such changes were reported after six screening trials. Prior to the experiment, participants were instructed how to hold chopsticks (one pair in each hand) and were allowed a brief practice. We instructed participants to keep their elbows on the table in front of them, positioning their hands ~30 cm in front of their face and ~35 cm apart. Participants maintained stable fixation by steadily gazing at a point halfway between their hands. After instructions, participants were dark adapted for ten minutes.

An experimental trial (Figure 1) started with participants using the chopsticks in each hand to pick up two cotton balls. A flash was emitted and participants verbally indicated the start of their afterimage, whereafter they dropped one of the two cotton balls. Participants described any perceived differences between the two sides: the Action Side from which the object was dropped versus the Stationary Side where nothing was dropped. To provide a measure of overall afterimage duration, participants indicated when the entire afterimage had faded back to complete darkness. After each trial participants had to pick the cotton balls back up in preparation for the following trial. In order to accomplish this we used a red laser pointer to help participants get their bearings in the dark. Since the wavelength of red light falls outside of the sensitivity range for retinal rods, this was a useful way to prepare for upcoming trials without disturbing dark adaptation.

Each participant performed 20 trials: a cotton ball was dropped from the chopsticks in the dominant and the non-dominant hand 10 times each, switching sides every 5 trials (counterbalanced across participants).

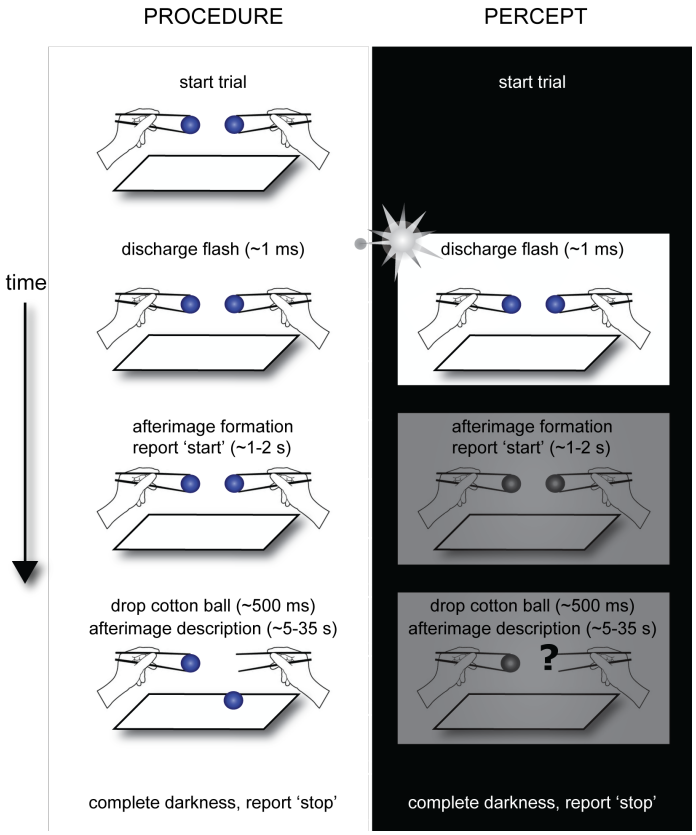


Figure 1. Trial sequence. An experimental trial began with participants sitting in complete darkness while holding two pairs of chopsticks, one in each hand, and a cotton ball between each pair of chopsticks. After discharge of the flash, participants were instructed to drop a cotton ball upon formation of the afterimage from either the dominant or the non-dominant hand. Participants then described any perceived differences between the objects in the afterimage, comparing the Action Side (where the cotton ball was dropped) with the Stationary Side (where the cotton ball was stationary).

After completion of the experimental trials, participants performed a skill test with chopsticks using their dominant hand. During part A (3 trials), participants moved 10 small, curved objects (uncooked macaroni, 8 mm in length) from one cup into another (diameter=8.4 cm; height=5.85 cm) over a distance of 20 cm. During part B (3 trials), participants moved 5 small square objects (standard dice, 15–mm cubes) over a distance of 20 cm, stacking them on top of each other. Direction of movement (left-to-right or right-to-left) was counterbalanced across trials, and the order of the two parts was counterbalanced across participants. We assume that faster performance indicates higher degree of skill, allowing the average trial duration to represent a measure of chopstick-skill. However, the ecological validity of part B was questionable, since it contained a spatial component and was prone to catastrophic errors in cases where the stacked objects would

tip over. This is reflected in the large variability of performance on part B (mean=43.44 sec; *SEM*=5.58 sec) compared to part A (mean=49.09 sec; *SEM*=2.74 sec). Since it wasn't possible to evaluate post-hoc which trials had involved catastrophic error(s), and to avoid classifying many participants as outliers, participant's chopstick proficiency was defined as the average trial duration on part A of our skill test. After outlier removal (99.3 coverage, *sd*≈2.7, removed N=1), average trial completion time was 47.75 seconds (*SEM*=2.44). Note that the main results of the research presented here are identical whether skill is defined based on the full test, or on part A alone.

After completing experiment 1, eight participants continued with chopstick training in a naturalistic setting. For a period of 4 weeks they ate at least one meal per day with chopsticks, using their dominant hand (15–28 meals per participant, mean=23.5 meals). Participants were tested again after training, exactly as in Experiment 1. Since training exclusively targeted the dominant hand, the non-dominant, non-trained hand served as a within-subject control.

Analyses. Reports on the appearance of objects in afterimages can vary widely between individuals (Davies, 1973a, 1973b). In this study, responses ranged from 'no perceived differences' to 'premature fading', 'transparent dimming', or even complete disappearance of one or several objects in the afterimage. We categorized responses as in Carlson et al. (2010), labeling reports (based on descriptions of the second-order objects) indicating greater fading on the Action Side vs. Stationary Side as positive responses. Two independent observers blindly rated every voice recording collected.

Due to the binary nature of our outcome measure (fading vs. no fading), and difficulties associated with modelling probability which has a restricted range of 0-1, we analyzed the data using logistic regression models in Stata (StataCorp., 2009). Specifically, we used a logistic random-intercept model, which allows fitting individual intercepts to participant's data to account for inter-individual differences in baseline fading experiences. Interactions were interpreted using simple slope analyses. In Experiment 1, skill was

modeled as a continuous between-subjects variable, and the hand used to perform the action was modeled as a categorical within-subjects variable. In Experiment 2 both training and hand were modeled as categorical within-subject variables.

Results

Experiment 1: Second-order fading, skill, and lifetime-built motor fluency. Second-order integration, as indicated by fading of a cotton ball held with chopsticks from the Action Side was experienced on 26% of experimental trials ($t=9.762$; $p<0.0001$, one-sided against zero). This finding uncovers the ability of human observers to rapidly integrate the chopsticks into the body representation, with the chopsticks providing rich cross-modal expectations about the effects of dropping the second-order cotton ball.

With regard to handedness, we find that at a mean skill level, the odds of experiencing fading are 46.4% higher when the cotton ball is dropped from the chopsticks in the dominant, as opposed to the non-dominant hand ($exp(B)=0.381$; $p=0.025$). Moreover, our data also revealed an interaction between hand and skill ($exp(B)=-0.031$; $p=0.004$). As depicted in Figure 2a, when dropping the cotton ball from the chopsticks in the non-dominant hand, skill level did not influence the amount of fading participants experienced ($exp(B)=0.001$; $p=0.948$). However, when using the dominant hand, more skilled (faster) participants perceived more fading of the cotton balls (larger log odds) than those participants whose skill with chopsticks was poorer. In fact, the odds of experiencing a cotton ball fading increased 0.03% for every second a participant was faster on the chopstick-skill test ($exp(B)=-0.029$; $p=0.035$). The model provided a good fit for the data (Wald Chi-Square(3)=16.33; $p=0.001$).

Note that the results depicted in Figure 2 are expressed in log odds of fading. The relationship between log odds and probability can be defined as:

$$\log \text{odds}(p) = \log\left(\frac{p}{(1\text{-}p)}\right)$$

Where *p* stands for the probability of fading. As a reference, log odds of zero correspond to a 50% chance of perceived fading, whereas log odds smaller than zero indicate a chance of fading which is less than 50%.

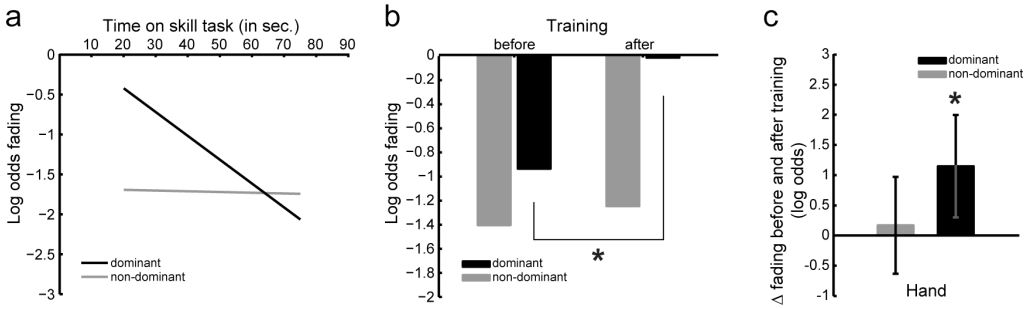


Figure 2. Second-order object integration is modulated by skill and training. Results from the skill and training experiments, expressed in log odds of fading. Note that 50% chance of fading corresponds to log odds of zero. Log odds < 0 indicate that chances of perceiving second-order object fading were smaller than 50% (in fact, overall chances of seeing a second-order object fade were 25.8%). **(A)** Skill modulates fading of a distal cotton ball dropped from a pair of chopsticks. When using the dominant hand (black line, depicting fixed-effects logistic regression) participants experienced more second-order fading when they were more skilled with the first-order tool (ergo they were faster on the skill test). For highly skilled participants using their dominant hand, the log odds of seeing cotton ball fading were closer to zero, which means that they were more likely to experience fading. No such relationship between skill and fading was found when participants used their non-dominant hand (grey line). **(B)** Training leads to improved long-term integration of a second-order object into the body representation. After extensive chopstick practice with the dominant hand, fading experiences of the cotton ball increased to log odds of almost zero (about half of all trials) for the trained hand only (black bars). No effect of training is found for the non-dominant hand (grey bar). Bars depict the (fixed-effects) logistic regression estimates of fading before and after training, with the interaction term included. **(C)** Each bar depicts the difference between the log odds of fading before an after training, for each hand separately. This plot also shows that training increases the log odds of fading for the trained dominant hand, but not for the untrained non-dominant hand. Error bars indicate the 95% confidence interval around the difference scores.

Experiment 2: Training and second-order fading. An assessment of the amount of fading experienced by our participants before and after training revealed that training

improved the odds of experiencing cotton ball fading by 76.36% (exp(B)=0.567; p=0.037; Figure 2b). This is a strong indication that modifications to the representation of one's own body can be strengthened over time by extensive tool practice. Independent of training, participant's odds of experiencing fading of the cotton ball were higher (138%) when the cotton ball was dropped from the chopsticks in the dominant, as opposed to the non-dominant hand (exp(B)=0.867; p=0.002; Figure 2b). This finding mirrors results from Experiment 1. The overall model predicted the data well (Wald Chi-Square(2)=13.75; p=0.001).

We expected improved long-term integration of the first-order tool for the trained (dominant) hand, but not for the untrained (non-dominant) hand. Hence, we also tested the impact of training on second-order object fading for both hands separately (Figures 2b and 2c). Participants' odds of experiencing cotton ball fading were 215.19% increased for the dominant hand after training (exp(B)=1.148; p=0.008). This finding holds true after a strict correction for multiple comparisons (p=0.016). In contrast, the non-dominant hand shows no differences before and after training (exp(B)=0.168; p=0.682).

Finally, a 2-way repeated measures ANOVA showed that the duration of the afterimage was not affected by training ($F_{(1,7)}$=1.788; p=0.223), nor the hand used to drop the cotton ball ($F_{(1,7)}$=0.853; p=0.387). Taken together, Experiments 1 and 2 demonstrate participants' ability to integrate both first-, as well as second-order objects into the body representation, whereby the success of second-order integration changes as a function of familiarity with the first-order tool.

Skill performance and individual differences. After the 4-week training, skill with chopsticks was significantly improved (t=2.745; p=0.029; Figure 3a). Interestingly, we also found a high correlation between the time spent on the skill task at baseline and skill improvement (r=0.921; p=0.001; Figure 3b). This implies that participants with low baseline-skill benefited most from training. A possible explanation for this could be near-ceiling performance of high-skilled individuals at baseline. Regression toward the mean is

a less likely interpretation, since participants who were slower at baseline got faster after training, whereas participants who were already fast at baseline did not really get slower. The reduction of variance after training also argues against a random redistribution of skill-test scores.

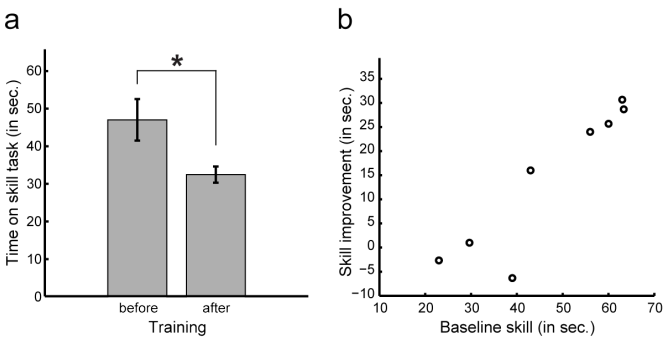


Figure 3. Skill with chopsticks before and after training. **(A)** Participants needed less time to perform an average skill trial after prolonged training with chopsticks (p=0.029) **(B)** Participants who were worse (slower) on the chopstick-skill test before training, appear to show the largest benefit of extensive tool practice after training. Skill improvement is defined as ($skill_{before}$ − $skill_{after}$) in seconds.

Large inter-individual differences are characteristic of the way changes in positive afterimages are perceived (Davies, 1973a, 1973b). Here we find that, despite large between-subject variability (0–90% fading experienced across participants), the amount of second-order object fading within individual participants is fairly stable. The frequency of fading experiences was correlated between the dominant and non-dominant hand (r=0.494; p=0.0003). Moreover, comparing the amount of fading before and after training explains 33.87% of variance (r=0.582; p=0.018), indicating stability over time. Thus, the amount of fading a person experienced could be partly explained by individual proneness to such fading.

Discussion

This study demonstrates that the representation of the body can be extended beyond first-order limitations to also include second-order objects held by a tool. Participants dropped cotton balls from a pair of chopsticks, and these cotton balls faded from a stable visual scene on approximately one-fourth of all occasions. This expands the domain in which

second-order fading has been demonstrated, since fading of "objects held by held objects" has not been observed before in situations where tool-use mainly relies on somatosensory and kinesthetic information from more distal parts of the hand.

We furthermore demonstrated that this fading effect was modulated by both skill and learning. Skillfulness with the first-order chopsticks actively influenced the amount of rapid integration found for the second-order cotton balls, as indexed by a higher degree of fading of these cotton balls in more-skilled participants. A slower buildup of integration was found after extensive training with the chopsticks, indicating an additional long-term component. Thus, skill and learning can interact with the extended representation of the body. No systematic changes in afterimage durations were uncovered across the various experimental conditions. This helps rule out observer bias, since we relied on subjective reports to determine both fading of imaged objects, as well as fading of the entire imaged scene. Since scene complexity was constant across conditions, the lack of variability in afterimage duration is in line with previous findings (Davies, 1973a).

After a month of practice, fading of cotton balls dropped from chopsticks in the trained (dominant) hand increased to 50% of observations. Fading of cotton balls from chopsticks in the untrained (non-dominant) hand remained unchanged, revealing the specificity of this training-induced modification of the body representation in our participants. Throughout the experiment, participants experienced more fading when using their more-skilled dominant hand (29.1%–32.9%), compared to their non-dominant hand (22.4%–26.3%). This difference might be explained by a general difference in lifetime built motor-fluency between the two hands; the dominant hand typically being the more practiced. Motor-fluency reflects motor ability in a more general sense, and can be considered independent from tool-specific skill, which is the type of skill people acquire through (extensive) experience with a specific tool. Our results show that at very low levels of skill, similar amounts of second-order fading were experienced irrespective of the hand used. General differences in motor fluency cannot fully account for this, indicating that

tool specific experiences (like those acquired during training) might prove integral for assimilation of external objects into an extended representation of the body.

Furthermore, the finding from Experiment 2 that training leads to more second-order object fading is an important one, since it allows us to draw an even more definitive distinction between general motor fluency and tool-specific skill. General motor fluency and tool-specific skill are probably often correlated, but there can be deviations between the two. To take an extreme example, a generally dexterous undergraduate student with no experience handling chopsticks might do better on the skill task from our experiment compared to a clumsy 90-year old subject who has eaten with chopsticks their whole life. Which person would see more fading? As this example demonstrates, general skill or dexterity might not necessarily mean that a person has more tool-specific experience. Thus, our first experiment does not provide an unequivocal answer to the question why people who do better on the skill test see more fading. Based on the results from Experiment 1, participants who saw more fading may have had more tool-specific skill with chopsticks, but it is also possible that these participants were simply less clumsy and more dexterous in general. The second experiment resolves this question, favoring an interpretation that stresses actual tool experience as a modulator of the fading effect.

Selective fading of an object from the afterimage has previously been considered evidence for integration of that object into the *body schema*, based on the underlying assumption that only items which are part of the body schema will fade (Carlson, et al., 2010). However, any region in the afterimage where conflict arises between vision and proprioception is susceptible to fading, and fading can be modulated by higher-order experiences such as sense of ownership (Hogendoorn, et al., 2009). Fading might therefore be more conservatively characterized as the resolution to a conflict between the visual afterimage and the expectation that forms on the basis of somatosensory and proprioceptive information – instead of proof that an object was integrated into the body schema. Thus, fading is a demonstration that we have a rich, context sensitive ability to

formulate cross-modal expectations about the behavior of external objects with which we interact.

The importance of training for improved tool-integration, as indexed by fading of second-order objects from the positive afterimage, could stem from a honing of a participant's ability to differentiate amongst subtle yet complex mechanical forces sensed through dynamic touch. If such sources of information are not clearly discernable prior to training, an increased sensitivity of the haptic perceptual system could be an important requirement for becoming a more fluent tool-user. We demonstrate here that with experience, a tool that was not originally part of the body becomes capable of providing sufficiently rich cross-modal expectations about the effects of tool actions, such that the tool essentially becomes incorporated in the body representation as an effector. By incorporating a tool (chopsticks), the consequences of motor actions with that tool (releasing a cotton ball) become better predictable. Thus, the ability to flexibly map movements and their consequences proves paramount to the integration of first- and higher-order objects into the body representation.

We have assumed here that this improved link between multimodal predictions and action consequences means that tool-use modifies our representation of peripersonal space. However, what can these results tell us about the body schema, or even the body image? Probing the *body schema* can be done via tasks involving proper tool-use, which has been defined as using a tool in a way that includes a causal interaction, and contact with, the object acted upon (Cardinali, et al., 2012). According to this definition, observers in our experiment were involved in actions (namely dropping a cotton ball from a pair of chopsticks) that tap into the unconscious body schema representation. Nevertheless, our observer's task was perceptual in nature, namely judging which of two cotton balls is more visible in an afterimage and giving a verbal response. Such a task does not directly explore kinematics; potential integration of the chopsticks and cotton ball into the body schema of our participants therefore remains only tentative.

Though perceptual in nature, our task also does not directly probe the way chopsticks and cotton ball are sensed by our participants, as would be required when investigating the *body image*. Instead, our study uses participant's perceptual reports as an indirect measure for an underlying system for action and tool-use. This interpretation (that our task does not probe the body image) is in line with previous research demonstrating that tools cannot be integrated into the body image (Cardinali, et al., 2011; Cardinali, et al., 2012). For example, grasping movements only affected the report of arm length when reports were made via a pointing movement towards a tactile stimulus on the arm (emphasizing the body schema), but not when reports and location on the arm were indicated verbally (emphasizing the body image) (Cardinali, et al., 2011).

Humans have uniquely adapted for the use of tools with a flexibility and versatility that far surpasses other primates (Davies, 1973a; Seed & Byrne, 2010). Non-human primates on the other hand demonstrate relatively rudimentary tool-usage with moderate levels of inferential causal reasoning (Fujita, Kuroshima, & Asai, 2003; Goodall, 1986; McGrew, 2010; Vaesen, 2012; Visalberghi, et al., 2009). Given these differences at the behavioral level, one could hypothesize that humans have evolved to naturally and rapidly assimilate first-order representations – thus having an innate capacity for tool-use (Maravita & Iriki, 2004; Peeters, et al., 2009; Vaesen, 2012). Conversely, non-human primates may require substantive tool training to initiate the appropriate brain changes, which could include reorganization of somatosensory and visual signals (Ishibashi, et al., 2002a; Ueno & Fujita, 1998) and the creation of novel neural connections (Hihara, et al., 2006; Ishibashi, et al., 2002b).

We have demonstrated here that humans can integrate not only first-, but also second-order representations, whereby the success of second-order integration changes as a function of familiarity with the first-order tool. Thus, when it comes to strengthening the feedback loop between tool perception and action consequences via training (thereby improving predictions for second-order extensions), the human situation might be akin to that of non-human primates – training is required to get the brain wired up for the task.

For the time being, these conclusions remain tentative, since the task differences between most human and monkey work are substantial, and different processes could be involved.

The work presented here aims to bridge two diverging directions in the current literature, one involving rapid integration suggested by human psychophysical experiments (Bruggeman, et al., 2013; Cardinali, et al., 2009; Carlson, et al., 2010), while the other is the much slower buildup of representations described in monkey physiology (Iriki, et al., 1996). Based on the novel findings presented here, we suggest that the ability to modify body representations to include external objects can happen continuously, without previously suggested discrete limitations in terms of time and space (Carlson, et al., 2010). We suggest that integration might happen anywhere along a temporal continuum: One might expect very rapid integration for highly familiar or intuitive extensions, but more limited –practice dependent– integration when items are unfamiliar. Additionally, such variations in integration speed are likely related to the amount of sensorimotor feedback provided by a tool, which can be improved via training. Also, we propose that the degree with which one can extend oneself into the environment is not constrained in an absolute sense: as the number of extensions increases, the probability of integration might drop, but such constraints could hypothetically be lifted given enough training. A more continuous view might similarly explain differences between species, with more dexterous tool-users hypothetically having more potential in terms of the possible number of extensions, or speed of integration. Future research could establish the exact limits of the capacity to incorporate objects not originally part of the body.

References

Berti, A., & Frassinetti, F. (2000). When far becomes near: remapping of space by tool use. *J Cogn Neurosci, 12*(3), 415-420.

Botvinick, M., & Cohen, J. (1998). Rubber hands 'feel' touch that eyes see. *Nature, 391*(6669), 756.

Bross, M. (2000). Emmert's law in the dark: active and passive proprioceptive effects on positive visual afterimages. *Perception 29*(11), 1385-1391.

Bruggeman, H., Kliman-Silver, C., Domini, F. & Song, J.H. (2013). Dynamic manipulation generates touch information that can modify vision. *Psychological Science*, *24*(6), 1063-1065.

Cardinali, L., Frassinetti, F., Brozzoli, C., Urquizar, C., Roy, A. C., & Farne, A. (2009). Tool-use induces morphological updating of the body schema. *Curr Biol, 19*(13), 1157.

Cardinali, L., Brozzoli, C., Urquizar, C., Salemme, R., Roy, A. C., & Farnè, A. (2011). When action is not enough: tool-use reveals tactile-dependent access to Body Schema. *Neuropsychologia, 49*(13), 3750-3757

Cardinali, L., Jacobs, S., Brozzoli, C., Frassinetti, F., Roy, A. C., & Farnè, A. (2012). Grab an object with a tool and change your body: tool-use-dependent changes of body representation for action. *Exp Brain Res, 218*(2), 259-271

Carlson, T. A., Alvarez, G., Wu, D.-A., & Verstraten, F. A. J. (2010). Rapid Assimilation of External Objects Into the Body Schema. *Psychological Science, 21*(7), 1000-1005.

Colby, C. L. (1998). Action-oriented spatial reference frames in cortex. *Neuron, 20*(1), 15-24.

Davies, P. (1973a). Effects of movements upon the appearance and duration of a prolonged visual afterimage: 1. Changes arising from the movement of a portion of the body incorporated in the afterimaged scene. *Perception, 2*, 147-153.

Davies, P. (1973b). Effects of movements upon the appearance and duration of a prolonged visual afterimage: 2. Changes arising from movement of the observer in relation to the previously afterimaged scene. *Perception, 2*, 155-160.

Farnè, A., Iriki, A., & Làdavas, E. (2005). Shaping multisensory action-space with tools: evidence from patients with cross-modal extinction. *Neuropsychologia, 43*(2), 238-248.

Fujita, K., Kuroshima, H., & Asai, S. (2003). How do tufted capuchin monkeys (Cebus apella) understand causality involved in tool use? *J Exp Psychol Anim Behav Process, 29*(3), 233-242.

Gibson, J. J. (1966). *The senses considered as perceptual systems.* Boston: Houghton Mifflin.

Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge, MA: Harvard University Press.

Goodwin, G. M., McCloskey, D. I., & Matthews, P. B. (1972). Proprioceptive illusions induced by muscle vibration: contribution by muscle spindles to perception? *Science, 175*(4028), 1382-1384.

Graziano, M., Yap, G., & Gross, C. (1994). Coding of visual space by premotor neurons. *Science, 266*(5187), 1054-1057.

Gregory, R. L., Wallace, J. G., & Campbell, F. W. (1959). Changes in the size and shape of visual after-images observed in complete darkness during changes of position in space. *The Quart J of Expt Psych 11*(1), 54-55.

Head, H., & Holmes, G. (1911). Sensory disturbances from cerebral lesions. *Brain, 34*(2-3), 102-254.

Hihara, S., Notoya, T., Tanaka, M., Ichinose, S., Ojima, H., Obayashi, S., et al. (2006). Extension of corticocortical afferents into the anterior bank of the intraparietal sulcus by tool-use training in adult monkeys. *Neuropsychologia, 44*(13), 2636-2646.

Hogendoorn, H., Kammers, M. P. M., Carlson, T. A., & Verstraten, F. A. J. (2009). Being in the dark about your hand: resolution of visuo-proprioceptive conflict by disowning visible limbs. *Neuropsychologia, 47*(13), 2698-2703.

Holmes, N. P. (2012). Does tool use extend peripersonal space? A review and re-analysis. *Exp Brain Res, 218*(2), 273-282.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport, 7*(14), 2325-2330.

Ishibashi, H., Hihara, S., Takahashi, M., Heike, T., Yokota, T., & Iriki, A. (2002a). Tool-use learning induces BDNF expression in a selective portion of monkey anterior parietal cortex. *Brain Res Mol Brain Res, 102*(1-2), 110-112.

Ishibashi, H., Hihara, S., Takahashi, M., Heike, T., Yokota, T., & Iriki, A. (2002b). Tool-use learning selectively induces expression of brain-derived neurotrophic factor, its receptor trkB, and neurotrophin 3 in the intraparietal multisensorycortex of monkeys. *Brain Res Cogn Brain Res, 14*(1), 3-9.

Kammers, M. P. M., Kootker, J. A., Hogendoorn, H., & Dijkerman, H. C. (2010). How many motoric body representations can we grasp? *Exp Brain Res, 202*(1), 203-212.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends Cogn Sci (Regul Ed), 8*(2), 79-86.

Maravita, A., Spence, C., & Driver, J. (2003). Multisensory integration and the body schema: close to hand and within reach. *Curr Biol, 13*(13), R531-539.

McGrew, W. C. (2010). Evolution. Chimpanzee technology. *Science, 328*(5978), 579-580.

Peeters, R., Simone, L., Nelissen, K., Fabbri-Destro, M., Vanduffel, W., Rizzolatti, G., & Orban, G. A. (2009). The Representation of Tool Use in Humans and Monkeys: Common and Uniquely Human Features. *Journal of Neuroscience 29*(37), 11523-11539.

Quallo, M. M., Price, C. J., Ueno, K., Asamizuya, T., Cheng, K., Lemon, R. N., et al. (2009). Gray and white matter changes associated with tool-use learning in macaque monkeys. *Proc Natl Acad Sci USA, 106*(43), 18379-18384.

Ritchie, J. B., & Carlson, T. (2010). Mirror, mirror, on the wall, is that even my hand at all? Changes in the afterimage of one's reflection in a mirror in response to bodily movement. *Neuropsychologia, 48*(5), 1495-1500.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science, 277*(5323), 190.

Seed, A., & Byrne, R. (2010). Animal Tool-Use. *Current Biology, 20*(23), R1032-R1039.

Tomasello, M., & Call, J. (1997). *Primate Cognition*. New York: Oxford University Press.

Turvey, M. T. (1996). Dynamic touch. *American Psychologist, 51*(11), 1134-1152.

Ueno, Y., & Fujita, K. (1998). Spontaneous Tool Use by a Tonkean Macaque (Macaca tonkeana). *Folia Primatol, 69*(5), 318-324.

Vaesen, K. (2012). The cognitive bases of human tool use. *Behav Brain Sci, 35*(4), 203-218.

Vignemont, F. (2010). Body schema and body image - pros and cons. *Neuropsychologia 48*(3), 669-680.

Visalberghi, E., Addessi, E., Truppa, V., Spagnoletti, N., Ottoni, E., Izar, P., et al. (2009). Selection of Effective Stone Tools by Wild Bearded Capuchin Monkeys. *Current Biology, 19*(3), 213-217.

Wagman, J. B., & Carello, C. (2001). Affordances and inertial constraints on tool use. *Ecological Psychology, 13*(3), 173-195.

Wolpert, D. M., Goodbody, S. J., & Husain, M. (1998). Maintaining internal representations: the role of the human superior parietal lobe. *Nat Neurosci, 1*(6), 529-533.

Wolpert, D. M., & Flanagan, J. R. (2010). *Curr Biol, 20*(11), 467-472.

Yamamoto, S., & Kitazawa, S. (2001). Sensation at the tips of invisible tools. *Nat Neurosci, 4*(10), 979-980.

Chapter 7

Summary and Conclusions

Summary of the main findings

In this thesis we used behavioral, computational, imaging and brain stimulation techniques to investigate the short-term retention of visual orientation. We also explored how the body's representation can be flexibly updated to include external objects such as tools. Here we will briefly summarize our findings, and discuss the main implications of our work.

Memories are imperfect, and even memory for a basic visual feature like orientation dwindles over time or due to interference by other visual stimuli. In Chapter 2 we demonstrated that internal mnemonic representations of orientation information are not immune against interference, and systematic biases emerge when a distractor is presented during the delay. These biases consist of an increase in the variability of the mnemonic representation, and an attraction of the mnemonic representation towards the orientation of the irrelevant distractor – constituting a sort of 'false memory' of the remembered orientation. Manipulating awareness and attention influenced these memory biases: Biases disappeared altogether when a distractor was presented outside of awareness, and when the distractor was attended there was a reduction of the attraction effect. While this work informs models of visual working memory and feedback related processing, it also raises some questions.

To take a step towards unraveling the mystery of *why* irrelevant information might be integrated, we proposed two computational models in Chapter 3. As a first pass these models try to explain the attraction effect under conditions of a task irrelevant distractor. The first of these models is a perceptual averaging account, which assumes that a mnemonic representation is nothing more than a weighted average of the target and distractor orientation. The distractor, being irrelevant, presumably carries less weight in the averaging process. From the findings in Chapter 2 it's obvious that such a model would fall short trying to generalize to other conditions. For example, it would not be able to explain the reduced attraction when the distractor is task relevant – the weighted

averaging model assumes that adding attention to the mix would increase the distractor weight, wielding a larger rather than a smaller influence during averaging.

The second model explains the attraction effect as a two-stage process. First a decision is made about an orientation relative to cardinal, and information incongruent with this decision is dismissed. This decision stage alone can explain the so-called 'default bias' found even in the absence of a distractor, which consists of a response bias away from the cardinal axes. Moreover, this decision stage could potentially explain the oblique effect as well, since more cropping of distributions closer to cardinal would reduce the variance of report. Second, the target and distractor representations undergo multiplicative integration, where the distractor is represented with more noise than the target. Both the perceptual averaging and the decision and integration model explained the attraction effect fairly well. The decision and integration model has great potential for predicting front-end perceptual orientation processes, yet it is more difficult to generalize it to processing of other visual features or higher-order stimuli.

Chapter 4 demonstrated the seemingly obvious, namely that very basic visual attributes, as well as more complex visual objects, cannot be maintained perfectly over time. First of all, experiments in this chapter confirmed the intuition that over time the quality of visual representations inadvertently diminish. As a general rule, memory quality suffers the sharpest falloff early during retention rather than late (Fahle & Harris, 1992; Vogels & Orban, 1986). Such decay has been tentatively described to occur in a logarithmic fashion (Laming & Sheiwiller, 1958; Lewy, 1895; Wolfe, 1886), though at delays longer than an hour decay asymptotes and becomes less than predicted by the logarithm (Laming & Scheiwiller, 1958). Indeed, logarithmic decay cannot hold mathematically, since performance would tend to infinity at a delay of zero, and lead to negative memory precision after a finite duration of time. It's therefore more likely that memories suffer from exponential decay, as many things known in nature. The way in which memories are forgotten says a lot about how they are maintained. Assuming that recurrent networks and attractor states are responsible for maintaining representations once they can no

longer be directly verified via sensorial input (discussed in detail in the General Introduction), would one assume increases in noise, or the sudden termination of memory representations over time? We demonstrated that the former appears the most prominent mechanism of decay, while the latter cannot be entirely excluded. These findings can help shape future models of memory maintenance by providing clear standards on how decay manifests in the behaving human brain.

Memories are susceptible to interference from other visual stimuli as well as the passive passing of time. Actively disrupting memories of visual orientation with TMS would therefore seem trivially easy as a next step. The importance of early visual cortex during the maintenance of precise representations (Ester, Anderson, Serences, & Awh, 2013; Harrison & Tong, 2009; Serences, Ester, Vogel, & Awh, 2009) inspired the work in Chapter 5. Here we saw that TMS over early visual cortex can influence memory, though contrary to an anticipated interference, we found that memory representations were sharpened for stimuli at the targeted retinotopic location. Retinotopic aspecific interference was found only when pulses were delivered at the tail end of encoding, alluding to possible global encoding strategies used by participants, and interactions between individual items in the display. Note that with any TMS experiment, results critically rely on the experimental design and the stimulation protocol, especially the latter poses a challenge to the experimenter, who is required to pick a sensible value from a parameter space that is infinite. Nevertheless, uncovering improvements in performance is not without merit, and resonates with theoretical frameworks that suggest that random variability in the brain can actually play a functional role (McDonnell & Abbot, 2009) If true, stochastic function could lead to the counterintuitive observation that inserting random noise enhances the representation of a signal.

As we have discussed, most of what the brain is doing is not rooted in direct physical input, but buried in the computations occurring between vast amounts of cortico-cortical connections. Maintaining a visual stimulus over a temporal delay is one example of such processing, discussed in great detail in this thesis. The second example we have addressed

in this thesis concerns the representation of an object that is not originally part of the body, such as a tool. When handling a tool with the purpose to achieve some behavioral goal – such as picking up a pencil to jot something down, or grabbing a spoon to eat a bit of soup – the tool immediately and automatically feels as part of the body. Of course the tools in these examples are sensed by the mechanoreceptors of the hand, yet, no receptors are existent on the tool itself. Besides a few points of contact, the brain has to infer information regarding the tool, such as its location in space or its material properties. This process of inference is complete for objects handled with tools, for which there is zero direct sensory information. In Chapter 6 we showed that, despite these difficulties, even objects held by tools can be rapidly integrated in the body's representation. In order to extend the body beyond directly touched 'first-order' objects to 'second-order' objects (which are objects held by tools) skill with the tool matters: More integration of second-order objects was found when people exhibited a higher degree of skill with the first-order tool.

Together, the work presented in this thesis shines some light on very basic processes performed by the brain on a day-to-day basis. Humans are constantly engaged in the short-term retention of visual information, relying on an indispensable working memory buffer to extend representation of relevant sensory input over temporal gaps, and protecting them from interfering inputs. Relevant sensory information can then be used to achieve cognitive, and ultimately behavioral goals. In the human world, behavioral goals are often expressed through the use of tools, be it the daily maneuvering of a bike, scooter, or car through traffic in order to get safely to work, or picking up greasy dim sum with a pair of chopsticks, trying not to look like an idiot by having it fall and splash sauce all over your clothes during a work dinner. All these mundane acts – remembering, moving, or avoiding social awkwardness – barely scratch the surface of the full depths of human ability. By studying the human brain, and sticking with it, we might ultimately acquire a larger picture about human nature. And that picture might turn out to be *us*, looking bewildered, not having a clue.

References

Ester, E.F., Anderson, D.E., Serences, J.T., & Awh, E. (2013). A neural measure of precision in visual working memory. *Journal of Cognitive Neuroscience*, *25*(5), 754–761.

Fahle, M., & Harris, J. P. (1992). Visual memory for vernier offsets. *Vision Research*, *32*(6), 1033–1042.

Harrison, S., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.

Laming, D., & Scheiwiller, P. (1985). Retention in perceptual memory: a review of models and data. *Perception & Psychophysics*, *37*(3), 189–197.

Lewy, W. (1895). Experimentelle Untersuchungen iiber das Gediichtnis. *Zeitschrift fUr Psychologie und Physiologie der Sinnesorgane, 8,* 231–292.

McDonnell, M. D., & Abbott, D. (2009). What Is Stochastic Resonance? Definitions, Misconceptions, Debates, and Its Relevance to Biology. *PLoS Computational Biology*, *5*(5), e1000348.

Serences, J.T., Ester, E.F., Vogel, E.K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*(2), 207–214.

Vogels, R., & Orban, G. A. (1986). Decision processes in visual discrimination of line orientation. *Journal of Experimental Psychology: Human Perception and Performance*, *12*(2), 115–132.

Wolfe, H.K. (1886). Untersuchungen uber das Tongedachtniss. *Philosophische Studien*, 3, 534–571.

Chapter 8

Knowledge valorization

Uniquely human

Us humans, we like to think we're special. Oh so very special – better – smarter – more complex – more *awesome* than all the other animals. We say we are *the only ones* who can use tools, but then discover that monkeys, crows, otters, elephants, dolphins, and even octopuses use tools all the same. We say we are *the only on*es who harbor skills of moral reasoning or cooperation, until it is pointed out to us the remarkable abilities of all sorts of other animals – again not just highlighting the ability of our fellow primates, but even elephants and the likes. So special are we that even dogs and monkeys have traveled to space, while you, reading this, most likely did not. We are *the only ones* who can feel bottomless compassion, and then groups of us take it upon ourselves to go out in the thick of the night to assault our fellow human beings, even if those others have fled a rape and destruction and war, and are merely seeking refuge. All of this we *are* and worse, far worse. We prey, hurt, attack, and kill. The weak, the innocent, the children – just like the other animals.

Yet, us humans, we are *the only ones* who seek knowledge, *purely for the sake of knowledge,* because we're *curious*. What madness is this, building sandcastles in the mud? None of the other animals would even for a second dream of carrying on anything quite so futile and fruitless. But wait… Our Quixote-like endeavors are actually *useful*. We see the invention of penicillin, a dramatic rise in the quality of life due to the discovery of electricity, an age of information. We now see the others around us – if we look up from our screens – also engrossed in their personal devices, avoiding each and every possibility to establish contact with us on a person-to-person kind of level. Spurred on by these great advances we now want more, more, and still more. As we like to buy, buy, and buy still more, and thus need something new to be out there for us to obtain. There is no question as to whether this will enhance our lives as it MOST DEFINITELY will, how could it not? So, we rally up our scientists to do more of *that thing they do*, you know, that *useful thing*. And we will call it *valorization*. Us humans, we were the only ones who seek knowledge, *purely for the sake of knowledge.* And now, we are unique no longer.

Valorizing Valorization

The Merriam Webster Dictionary defines the word valorize \ˈva-lə-ˌrīz\ as follows:

(1) To enhance or try to enhance the price, value, or status of by organized and usually governmental action <using subsidies to valorize coffee>
(2) To assign value or merit to: to validate.

The Maastricht University Valorisation Center states that:

Valorisation is the process of creating value from knowledge, by making this knowledge available and suitable for economic and social exploitation and to translate this knowledge into products, services, processes and new business.

Thus, the Valorization of Knowledge is the process of trying to find how knowledge obtained via scientific research could benefit society as a whole, or be profitable. By logical deduction, the Valorization of Valorization (also commonly known as 'meta-Valorization') is the process of trying to find how Valorization of Knowledge could benefit society as a whole, or be profitable. Let's start by a simple calculation: In 2012 about 4000 PhD students obtained their degree in the Netherland, and this number has been rising[1]. To air of the conservative side we will assume that anno 2015 this number is still the same. Similarly, let's assume that PhD students are efficient little workers, and that the average student can write a Valorization chapter in about an hour (my office mate took 2 days, but was by his own account "very relaxed" throughout these proceedings). Now, let's assume that all PhD students in the Netherlands are forced to write a Valorization chapter for their thesis, we get this simple sum: 4000 students x 1 hour per student = 4000 hours of work. A PhD student costs the University about 40 euros an hour[2], which at 4000 hours of work adds up to 160.000 euros. Of course, these 4000 Valorization chapters are checked by roughly 4000 supervisors, and people like the so-called "Valorization Officers". Presumably, these guys cost even more per hour, but OK, let's not calculate them in for now. How many PhD salaries might be paid from 160.000 euros? About two years worth is how many.

1. Source: The Dutch Central Bureau of Statistics
2. Source: Kamil Uludag, former department head Cognitive Neuroscience at Maastricht University

Aside from the mind wrenching pain it casts upon PhD students, especially those in fields of fundamental sciences, to write a Valorization chapter; Aside from the fundamentally deceptive crap they jot down in order to appease the requirement, stuff they don't really mean, believe, can justify or even *know*; Aside from the benefits to society if –instead– these students were out in the sunshine smiling at everyone and dancing in the streets because they were excused from having to write a Valorization chapter. In order to Valorize Valorization, we might want to make sure these 160.000 or so euros are not better spent on healthcare, education, helping refugees, planting trees, or saving white Rhino.

Yes, these are only rough calculations, and yes I am sketching a caricature that lacks some nuance perhaps. But are we really sure that what we are doing for the benefit of society, benefits society? And that what we're doing to make a profit makes a profit? I would cautiously suggest a Valorization panel to valorize the Valorization (a meta-Valorization panel if you will) just to make sure.

As an aside: The meaning of the word 'valorization' as used here at our University is altered somewhat from its generally accustomed usage in the English language. Of course, this is to cover its new meaning of creating value from knowledge. So I imagine a not-so-distant future in which diaspora from our University end up in foreign places that are not yet accustomed to this alteration. At a few, perhaps many, occasions the word will be dropped, and initially these diaspora will look like dunces. Then, the rest of the world might catch on, and all creativity and spontaneity will be sucked from science until she's left deflated and unsatisfied. Humans: the only ones who seek knowledge, *purely for the sake of knowledge.* This, to my mind, is what we have to look forward to when each and every budding scientist is forced to write a valorization paragraph – brownie points with the government, and all our alumni looking like dunces, en route to slay science.

A single neuron perspective

Imagine: you are a single neuron, a spiny stellate cell for example. Yes. Imagine being a spiny stellate cell – it will help you slip into character with less effort. OK. You are an awesome mighty spiny stellate cell, and you live somewhere amidst early sensory cortex, though you – of course – are blissfully unaware of this fact. You spike sometimes, that's your job. You gobble up some glucose, you hang with some astrocytes, and you listen to a great many other neurons and the tales they found worthy of relating to you. You listen carefully and you consider the evidence. You're on the job. What is the likelihood that what the others are spiking on about is actually happening? Yes? No? What should you do? A threshold is reached and you just do it, you spike. POW. Damn that felt right, a bit exhausting, you need a quick recharge, maybe a smoke.

Then I ask you: "Single spiny stellate cell, how is your spiking justified? What value is created by your knowledge?" You might reflect and realize you have no clue, you just do your thing because you love doing it, and you sink into a deep depression realizing that if you died tomorrow the brain would go on as usual without as much as a blip. Your value is essentially zero. Of course, if all neurons were eliminated the story would be different. Altogether, the neurons in all their complexness decide whether to watch one more video of a cat jumping into Many Too Small Boxes[1] or to continue writing this darned chapter. But quite frankly, the brain is *complex as shit*, and trying to figure out what the value is of each individual neuron borders on madness.

Yet, this is what brain scientists do, for years on end. Entire lives are dedicated to this mad pursuit just because we can. What we learn incrementally and veer off track, and by the end of it we might all prove to be wrong. Now, as a single scientist I sit somewhere amidst other scientists studying sensory cortex, for the most part blissfully unaware of the other sciences. I publish sometimes, that's my job. I gobble up some of Bandito's soups, and hang with some students. I go to conferences and listen to a great many other scientists

1. https://en.wikipedia.org/wiki/Maru_%28cat%29

and the tales they found worthy of relating to me. I listen carefully and you consider the evidence. I'm on the job.

If we want someone to be able to tell my value, or how my knowledge can create value, we might want to consider training an army of scientist-scientists, to study us for years on end, dedicate entire lives. Because, quite frankly, the brain sciences are *complex as shit*.

So. Just let the single neuron spike, as he *cannot earnestly answer* my question. When forced to, he might try to Obey and Please; he might give you likelihoods of nearby events, he might confabulate those from the faraway brain. If he is detecting an orientation in his receptive field he might say that it is part of a junction, that is part of a letter, that is part of a word, that is part of a sentence, that is part of a story that has some kind of value or meaning. That sentence could be part of an important document stating once and for all the future of mankind – a resolve to end all hunger, war, inequality, suffering. We would like that, because wishful thinking of the hopeless naiveté seem to reign supreme. But you know what? The orientation might not be part of a letter at all! It might be part of some obscene graffiti inside a dirty subway cart, or part of what you see written on your face in magic marker after you wake up from an accidental nap amongst your college mates. The point: there are infinite degrees of freedom here. As a single spiny stellate cell jammed in some obscure visual pathway only coding for the far periphery of space you might feel like you have quite the overview of what is going on, but darned, what kind of a clue do you have really? – I imagine that's what particle physicist feel like *all the time*.

How my research will create world peace

Reading valorization chapters of other recent graduates can prove inspiring. Shiny crystal balls, held up to the light. You know what? My research can create world peace and happiness for all!

Scenario 1. The media will pick up my research, and the misguided scientific writing in the popular press mightily impresses everyone. So much so, that I am offered a position in Government. In this position, I manage to grab full power, aided by the army, and I drastically change the monetary system, curb needless (consumer) spending, talk mindfully to the other Leaders in Government, and execute many more such interventions. World peace!

Scenario 2. My code – which as a postdoc I will of course share publically to increase my chances of finding a faculty position – will go viral. Its popularity results in its becoming integrated into new software everywhere, and I become rich and buy a helicopter. However, the code is so buggy that these new products all break down. The consequences are reminiscent the anticipated destruction from Y2K – and even worse! The world falls into total turmoil, and people can no longer rip each other to shreds on Internet forums. Altogether we build a new society from the roots up. World peace!

Scenario 3. One day, my paper on biased memory for visual orientation lands on the desk of a brilliant young researcher studying zebra fish. Reading the paper she gets so bored that she falls asleep, hitting her head on the desk very hard. Water splashes out of the fish tank, onto some papers, staining the ink. Upon awakening, the brilliant young researcher grabs her papers and runs off to a meeting. A dreamy colleague suddenly notices the stain and is reminded of this mega virus she studied a while back. A sudden vision overwhelms her, and back in the lab she uncovers a mega virus of great therapeutic potential. She ends up curing her colleague's concussion, and virtually every other mental illness within her lifetime. World peace!

Scenario 4. Some giant insight of great philosophical value and importance occurs (see Figure 1). An immense tolerance settles in like a soft summer breeze. World peace!

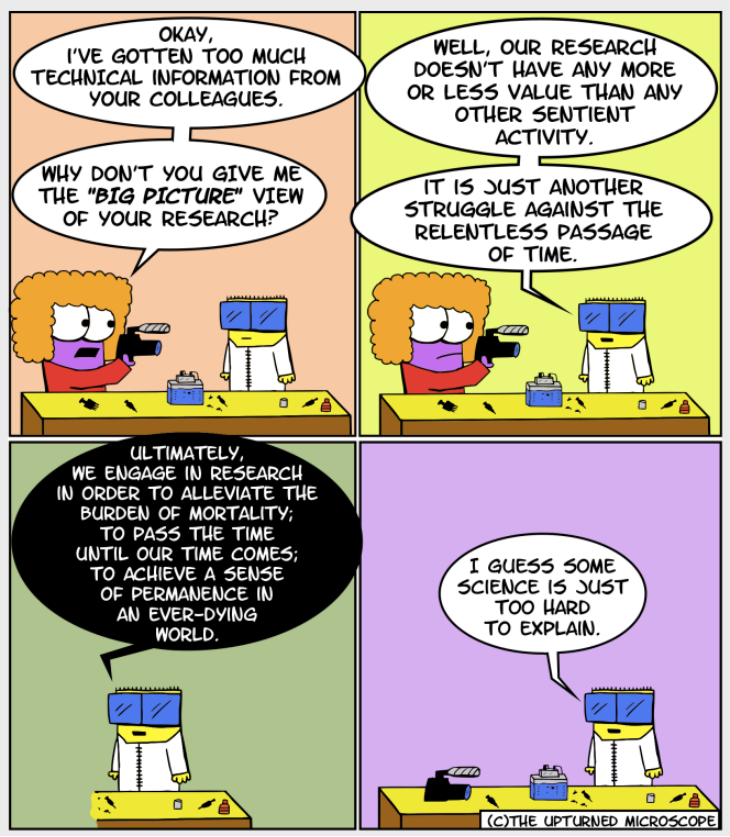


Figure 1. 'nough said

Reprinted with permission: Nik Papageorgiou – The Upturned Microscope

A cautionary tale of monster bugs

The following section is a cautionary tale for all those researchers out there whose work, like mine, depends critically on adequate coding skills. And whose work, like mine, may or may not ultimately result in world peace. My troubles came from writing some serious bugs into my experimental code, which I want to relate here in some detail. Specifically, there have been three that haunted me at night. While each bug is unique, and the variety of bugs that one can write into code seemingly endless, here follow three lessons, of bugs

not to repeat. (1) When you start using a pc to run experiments from, and when that pc – for some inexplicable reason – only wants to run your code the *first time* you run it after opening Matlab, MAKE SURE TO USE A RANDOM SEED. Computers are *deterministic*, and without a random seed (based on the clock for example) the machine will give you 'random' numbers that are *identical* every single time you restart your Matlab session. OK, good. (2) Don't overly complicate things by rotating the positions of your stimuli on every trial, it doesn't really matter that you want your vector of random orientations to move along with the location you probe your target at. If you have planted your seed atop your code just let the random numbers be and don't mess with them, or else… (Figure 2).

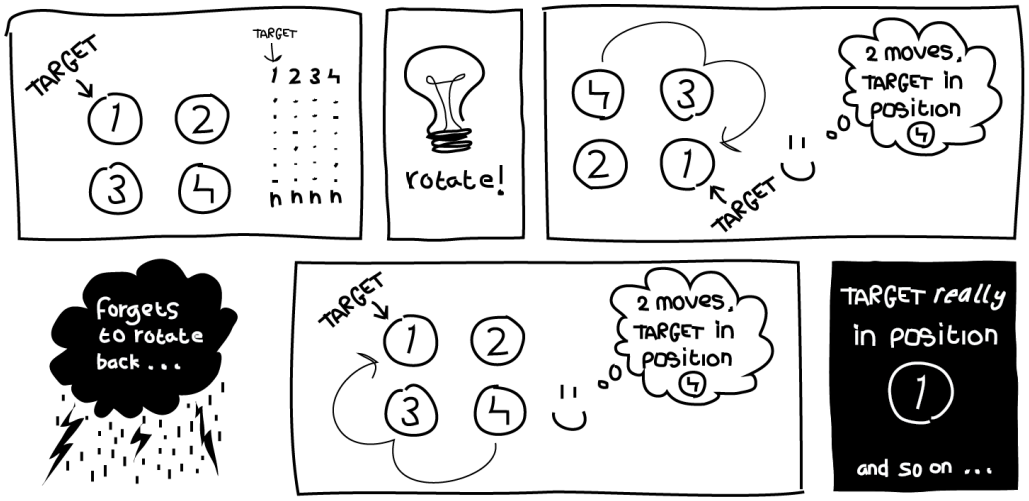


Figure 2. ♪ Ninety-nine bugs in my code at this point, ninety-nine bugs in my code. Prepare to frown, rotate back around, ninety-eight bugs in my code at this point. ♪

And last but not least (3) try to avoid separating your code into two chunks the day before your first scan. So what, your code doesn't run on the much slower experimental machine, fix *that*. Don't port the part generating your stimuli to another script to run on a faster machine but *leave the random seed outside of the run loop where it is needed*!

How random is life? Not random enough when you're writing code.

In all earnest

All joking aside – do the sciences have an immediate and foreseeable value? With respect to applied sciences; yes of course. Might advances to the benefit of society or for profit be achieved based on work done in the fundamental sciences? Absolutely, though I imagine purely by accident and in unforeseen manners. Might having PhD graduates write Valorization chapters into their theses create some kind of value (besides the shame they feel over those last couple of pages hidden in there following the real fruits of their hard labor over the past however-many years)? In very specific cases, sure, one could see how it might. Can young scientists foresee the future, or predict a chain of causality of two, three, or thousands of links into the future? Most of the time, *no*, most of the time we seem to be dealing with something rather entropic and chaotic. When a butterfly flaps its wings over the Atlantic… We can of course force PhD students to indulge in visions of a potential self-importance and bathe in delusions of grandeur on our way to turning a profit. Hell, maybe even my work *really matters* as a portion of humanity's collective knowledge, and one day it will *make me rich*. But let's not forget how small we really are, how single-neuron-small, how cog-in-a-ginormous-wheel small. If we want our science to keep creating 'lucky accidents' that we can all benefit from – penicillin, electricity – it needs room for such accidents; room for aimless venturing down unknown paths; for true discovery; for seeking knowledge *purely for the sake of knowledge*.

Acknowledgements

Science can feel lonely sometimes, when you're cocooned inside your own thoughts for hours on end, layers deep in code, concepts, or quest for meaning. But in truth, at least for me, the lonely scientist is wound tightly into a web of others just as absorbed by exploring the boundaries of what is known (or of what *they* know). These others inspire and enable growth, and are instrumental to shaping young scientist's minds. There is probably no right way to properly acknowledge or even know their exact contributions – for it is unclear where their input ends and that of the lonely scientist begins.

The first time I thought of myself as a scientist (granted, a bumbling one) was at Vanderbilt University in the lab of Frank Tong. Frank, as you're probably aware, you are my 'patient zero' – infecting me with passion for rigorous scientific practice, and want for an academic life. My aim is to live up to the standards you have imprinted onto my mind. Moreover, my time in the lab embedded me soundly in a community of talented scientists and generally great people who have played no small part in my life over the past years.

Janneke Jehee for example, a postdoc in the Tong lab back then, and now my co-supervisor – and incidentally one of the most amazing and driven people I know. Janneke, I admire you both as a scientist and friend, and cherish all those hours you listened to me and helped me navigate life and career. And there were others, too, Jan Brascamp and Tomas Knapen, also both at Vanderbilt (Blake lab) around that time, and both tangentially involved in some of my PhD projects. Jan, I still feel comfortable asking you for input when I see you on Skype, which means a lot to me. Tomas, remember how we wanted to show that visual imagery has a higher resolution than actual perception? I still have the 4 consecutive experimental scripts ready to go, just say the word. Back to the Tong lab: Mike Pratte, you're a total manwich (and so are you, Tim Kietzmann) and

please stop ‘not smoking’ when drunk. Joel, I still blame you for my bad coding practices, as yours were the first I got acquainted with. Jascha, thanks for having that couch, our many coffee runs, and your fascination with my past life that allowed me to feel like it was still a part of me. Eli, the quintessential postdoc to my mind, we bonded over our mutual lab frustrations and more general life rants. And thanks to all other lab members past and present.

No Sam Ling, I’m not forgetting you, but you’re a bit special, as you know, so here is a paragraph dedicated entirely to you. Back in my Tong lab days I got to know you as the guy that obsessively checked in on the mini-print of my first VSS poster – contagious laughter echoing down the hallways as an advance warning of your approach. In more recent years you have secretly acted as my daily supervisor. Your selfless and patient input has been tremendous, both in quality and quantity. Maybe I will stalk you on Skype forever. Maybe I will vicariously make your life bossier through Ilona. In any event, you are beyond a doubt one of my favorite people in the world. Sam for supervisor!

So Frank, let’s just say I’m not only happy for all your continued input on the projects we still work on together, but also for giving me all of this – setting me up for success before I even began my life as a PhD student.

The next phase: Maastricht, where all the work in this thesis was accomplished, and where Alexander Sack was the guy who believed in me and brought me on board. Alex, early on I thought of you as a sort of manager, skillfully overseeing the highly diverse output from your highly diverse group. Now I know better: You are an enabler! You’ve allowed me all the freedom in the world to pursue what I wanted, when I wanted it. And not just that, you have *sponsored* and *supported* all of this madness (to all participants who had their feet frozen off recently: you have this guy to thank for that). Although some of these scientific adventures turned out to be dead ends, you have nevertheless granted me the invaluable experience of running my own course. All the while, you have been highly attuned to the factors needed to succeed, and it’s absolutely refreshing to have such

remarkably fast turnaround on papers (of which we should soon publish many more, my turnaround is considerably worse than yours I admit).

When I joined the TMS lab (we call ourselves the TMS lab, why should we not be seen as the TMS lab?) I was welcomed by people like Teresa, Tom (hobby: make fun of Tom), Felix, Nina, Christianne, Joel, and Vincent. Vincent, you have been instrumental to guiding me through my early days here in Maastricht. You taught me how to use TMS (Chapter 5), and apparently I've managed to manage you to a point where you now explicitly ask me to "stalk me about it, I am a bit swamped with teaching at the moment". Careful what you wish for Vincent… I have your number on speed dial. In the interim, many new faces have appeared: Jeanette, Franzi, Sanne, Helen, Dennis, Geraldine, Lukas. And Tahnée: Our project almost made it into this thesis, now let's try and get it into yours (if only for the nice figures). It's sad that those countless hours of comical conversation, sitting in complete darkness (zapping people's brains, sometimes by accident) are over.

Speaking of collecting data – I can't claim to have done much of that by myself. Instead I have some great people to thank in this regard: Caroline Benjamins, every 2nd year research-practical student I've ever worked with, Ilona Bloem, Klaudia Ambroziak, Hanane Bittich, and Jana Devos. You guys are amazing, and it's quite obvious how I could not have done this without you. Ilona, I cannot believe my luck having had you as one of my students in the research practical. You have proven to be exceptional again and again, I think at some point you managed all the data collection of the work in Chapter 2, and you even piloted our dichoptic paradigm while I was out of the country. Your talent will take you wherever you want to go, and I feel much less worried about Sam knowing you are there to make sure he'll continue to be surrounded by *assertive* Dutch females.

Peter De Weerd, the notorious supervisor that I was told to fear. Instead, you've brought me friendship and collegiality. My experience working with you is that you are incredibly supportive and encouraging, and I'm honored by the seemingly atypical amount of trust that you have in my capabilities as a scientist.

Ruben van Bergen, graduate student in Janneke's lab, you may be *unbelievably young*, but I have actually learned a lot from working with you. My project at the Donders may not have worked out, but you are definitely one of the things I am keeping as a memento of good times. I am also keeping your code. We went to a castle, "Rauischholzhausen", this one time for a summer school on Visual Neuroscience. The castle, the people, the circumstances – it was a turning point for me, but a good kind of turning point, and I want to acknowledge *everyone* who was there. Pascal Mamassian in particular, because your talk at the summer school encouraged me to show you my poster, which then somehow led to the conception of Chapter 3. What I also find most astonishing is that after meeting me this one time you took the gamble of offering me a postdoc position, one I still wish I could have accepted. But even more baffling is your ability to teach: It was after a mere 3 days in Paris that I learned enough to do complete Chapter 3 two weeks later. It felt exhilarating learning so much so quickly, and I will remain in awe indefinitely. Oh, and please don't tell John Serences about that math thing we talked about over lunch with Sabrina, for at least another year or so.

Youngeun Park (another Tong lab treasure), you picked up on some work that later became the foundation of Chapter 4. Though you may appear shy at first, I've learned not to underestimate you, as you have a beautifully sharp mind and keep me on my toes. I always look forward to your e-mails and am even happier when they contain your comments and ideas, or edited manuscripts.

Admittedly, Chapter 6 is the odd chapter out. One night during VSS I was sitting on a couch at the Kanwisher party with this guy Daw-An Wu. From our interaction I mostly remember tales of fruit flies zipping through virtual environments. Daw-An, you also told me another thing that stuck, which was that a PhD should be about doing absolutely *nothing* but taking the time to read widely and follow where your interests take you. Inebriated or not, it's unique advice that I've not heard repeated since, but it has inspired some truly off the wall projects. Low and behold, several years after having accidentally landed next to you on a random couch we published our beautifully silly chopsticks paper,

aided by an overzealous group of 2nd year Psychology minions (yes Ilona, it's you again). This paper owes its existence to both of you – a dubious honor.

Now, science is not all about science and the people that science is done with. It's also about the times in between, and about digressing into conversations about whose Google algorithm is best trained *not* to yield pictures of girls in bikini when entering 'Crimea'; whether or not baby pee is actually sterile; who must go steal the balloons from the University event downstairs (making our office stink of balloons for weeks). Yes, Jan Zimmerman and João Correia, I am talking about you. Jan, I miss you, and I'm sorry for not choosing New York over sunny California. Ha. NOT. João, our friendship has been a complete one – we fought like little girls, and laughed about chili peppers with your grandma over a wonderful Portuguese dinner. New PhD students nowadays have it better than we did 5-or-so years back, and I think you are part of that equation – for being a sort of social glue; for fostering an environment that is interactive and content driven; by your excitement about the Big Discoveries that are just around the corner. Sri, I am so happy for you, and your soon-to-be legitimate reason to close all the blinds.

Right outside of the office doors, or actually loitering on the threshold (Gojko, how will I ever replace your 'always smiling' face), are many others. Like Anna, Mehrdad, Kamil (pizza party!), or Lars (why do we only ever talk when drinking?). Riny and Christl, thanks for all your support and letting me 'herumlungern' from time to time. Henna, thanks for being my friend. And previously also Marin, Matteo, Alexandros, Britta, and probably others who I've now forgotten – as soon I will be too. Marin, you are not forgotten but face a far worse fate, as you're my best friend in the whole of Maastricht, and now even paranimf to this thesis and defense. Our closeness was reflected back at me by Penelope this morning, when she was 'on the phone' with both you and Bram ("Hallooo!? Met Marin - ja - met Bram - ja hallo Marin"). My other paranimf (or should I say imp) Arianne, it's sort of crazy the different worlds that collided for us to become friends, and for us to remain friends ever since sultry India and those wildly decadent times. What now? Are we all grown up? Are we… old?

Boyfriend, while the things I did here are of little interest to you content-wise, you have suffered through some serious inconvenience in order for them to happen, and I am not referring only to your trusted proof reading skills. You may call me Dr. Girlfriend now, and maybe I'll do the voice. Penelope, if you are reading this some day: Maybe it's time to go do something more fun, by now I'm pretty sure the science in here is no longer very relevant.

Mijn lieve familie (mama, papa, andere papa, Irene, Jan, Shirley) waarbij ik groot heb mogen worden, maar altijd klein bij blijf: Ik ben blij met jullie allemaal. Voor mijn middelbare school vriendinnetjes geldt hetzelfde (Es, Niet, Lil, Val, San, Suus), evenals voor Abe, Fleur, Linda, en Roos.

Maastricht zou niet hetzelfde zijn zonder onze buurt aan de Bosscherweg – onze 'achterbuurt'. Erik, Marijn, en Niki, bedankt voor alle zomerse barbecues, wandelingen, theedrink sessies, honden uitwisselingen, ritjes naar het station en noem maar op. Jeroen en Laura, want wat zal het akelig stil zijn zonder al het geklus zo meteen, en ik ben vreselijk benieuwd naar het eind resultaat! Jan en Irma, jullie hebben mij en Penelope het prachtige dagelijkse avontuur gegeven van de zoektocht door de snoeptuin, die altijd weer een gezicht besmeurd met bramen, frambozen, druiven, bessen, of pruimen op mocht leveren. Dre en Lineke, Ray en Sanne, en alle patatjes, ijsjes, glijbanen en trampolines op de boerderij – onze buurt is de beste buurt van de *hele* stad.

En natuurlijk Coen en Petra: Het is niet eenvoudig om in een paar zinnen te vatten wat jullie voor mij betekenen, dat zal ik niet eens proberen. Wat ik wel weet is dat ik jullie in San Diego vreselijk ga missen, en dat ik nog steeds plannen aan het smeden ben om jullie met huis en al op een vrachtschip te laten verslepen die kant op. Natuurlijk weet ik dat het mijn eigen schuld is, omdat ik wel van een uitdaging houd, omdat ik dit zelf zo *wil*. Maar dat maakt het er nog niet gemakkelijk op, en neemt zeer zeker niet weg dat het eigenlijk helemaal niet leuk is om te vertrekken. Deze these is ook aan jullie te danken en al jullie

steun gedurende de afgelopen jaren. Ik hoop jullie nog heel erg vaak te zien, ookal zal dit per definitie nooit vaak genoeg zijn.

So there you have it, I acknowledge a whole bunch of people, every single one seemingly more special than the next. It reminds me of a recurring conversation in our office, where Jan would always comment on someone by saying "(s)he is sooo smart!", and João would always reply with "Yes, another brilliant person, according to you everyone is brilliant, name me one person who is not?!"). But indeed… All the people I have mentioned here, and all the ones who I may have forgotten, are quite brilliant indeed.

Publications

Rademaker, R.L., Bloem, I.M., De Weerd, P., & Sack, A.T. (2015). The impact of interference on working memory for visual orientation. *Journal of Experimental Psychology: Human Perception and Performance.* Advance online publication

Rademaker, R.L., Wu D-A, Bloem, I.M. & Sack, A.T. (2014). Intensive tool-practice and skillfulness facilitate the extension of body representations in humans. *Neuropsychologia* 56: 196–203

Rademaker, R.L., Tredway, C., & Tong, F (2012). Introspective judgments predict the precision and likelihood of successful maintenance of visual working memory. *Journal of Vision* 12(13) article 21: 1–13

Rademaker, R.L. & Pearson, J., (2012). Training visual imagery: Improvements of metacognition, but not imagery strength. *Frontiers in Perception Science* 3(224): 1–11

Pearson, J., **Rademaker, R.L.** & Tong, F (2011). Evaluating the mind's eye: The metacognition of visual imagery. *Psychological Science* 22: 1535–1542

Curriculum Vitae

At the unlikely hour of 6.17am on a Monday morning November 15th Rosanne Lynn Rademaker saw light for the first time in The Hague, The Netherlands, and a career as a vision scientist began. Playful at first, as a preschooler uncovering Troxler fading while lying on the bed staring up at the ceiling in the dim hours of the day, or 'seeing through her finger' when holding it close to her face and staring past in into the distance: both eyes open *invisible*! One eye closed *visible*! After these remarkable discoveries that took place while Rosanne lived in Zoetermeer, she moved across the Netherlands to the island of Texel at the age of 7. Here she finished primary school as well as High school (VWO, University Preparatory Level). Another move across the country ensued, to Groningen this time for a Bachelors in Psychology. Rosanne started modeling (the fashion, not the computational kind) and finished her degree successfully, though with some delay. After more time traversing the world in high-party style she returned to the country of her birth to start a Research Master program in Neuropsychology, at the University of Maastricht. Her second-year internship brought her to Vanderbilt University, where she ended up falling for science, the wonderful lab, as well as a cowboy-style man with wild curls, and she ended up staying for a period of almost two years. She moved back to Maastricht, cowboy in tow, and started her PhD program at the Cognitive Neuroscience department. This era is now also coming to a close, and a doctorate degree as well as a two-year-old richer, three will now move on to San Diego, USA, where Rosanne will start a postdoctoral research position at the University of California San Diego. To be continued.

“Penelope is lief, en mama is de baas, en héél groot, en héél dik”

– Penelope-ism November 2015.