IMAGES OF THE MIND

INVESTIGATING THE DYNAMIC INTERPLAY BETWEEN INTERNAL REPRESENTATIONS IN VISUAL CORTEX AND THE STRUCTURED EXTERNAL WORLD

MAËLLE LEREBOURG

RADBOUD
UNIVERSITY
PRESS

Radboud
Dissertation
Series

DONDERS
SERIES

Images of the Mind

Investigating the dynamic interplay between internal representations in visual cortex and the structured external world

Maëlle Lerebourg

Images of the Mind: Investigating the dynamic interplay between internal representations in visual cortex and the structured external world
Maëlle Lerebourg

Radboud Dissertation Series
ISSN: 2950-2772 (Online); 2950-2780 (Print)

Published by RADBOUD UNIVERSITY PRESS
Postbus 9100, 6500 HA Nijmegen, The Netherlands
www.radbouduniversitypress.nl

Design: Proefschrift AIO | Annelies Lips
Cover: Proefschrift AIO | Guntra Laivacuma
Printing: DPN Rikken/Pumbo

ISBN: 9789465151519
DOI: 10.54195/9789465151519

RADBOUD
UNIVERSITY
PRESS

Images of the Mind

Investigating the dynamic interplay between internal representations in visual cortex and the structured external world

Proefschrift ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de rector magnificus prof. dr. J.M. Sanders,
volgens besluit van het college voor promoties
in het openbaar te verdedigen op
dinsdag 28 oktober 2025
om 16.30 uur precies

door

Maëlle Silia Kalinka Lerebourg
geboren op 27 augustus 1995
te Berlijn

Promotoren:
Prof. dr. M.V. Peelen
Prof. dr. F.P. de Lange

Manuscriptcommissie:
Prof. dr. I. Toni
Prof. dr. J.J. Geng
Dr. N. Dijkstra

Images of the Mind

Investigating the dynamic interplay between internal representations in visual cortex and the structured external world

Dissertation to obtain the degree of doctor
from Radboud University Nijmegen
on the authority of the Rector Magnificus prof. dr. J.M. Sanders,
according to the decision of the Doctorate Board
to be defended in public on
Tuesday, October 28, 2025
at 16:30

by

Maëlle Silia Kalinka Lerebourg
born on August 27, 1995
in Berlin

PhD supervisors:
Prof. dr. M.V. Peelen
Prof. dr. F.P. de Lange

Manuscript Committee:
Prof. dr. I. Toni
Prof. dr. J.J. Geng
Dr. N. Dijkstra

Table of contents

References

Appendices

Chapter 1

General Introduction

As primates, our eyes are our main window into the outside world, allowing us to locate relevant objects and beings around us and to interact with them. A large part of our brain is therefore dedicated to process visual scenes, and we can often extract a vast amount of visual information, relating to the objects within them and their relations, at a single glance. As visual neuroscientists, we may become acutely aware of the challenges faced by our visual system, but as “users” we perceive a seamless and rich external world, mostly without devoting much thought or conscious effort to it. I will begin this introduction by outlining the difficulties of object recognition in cluttered scenes, focusing on competitive interactions between object representations in visual cortex. Further on, I will discuss how competition can be resolved by top-down biasing signals, the nature of these preparatory attentional templates and their role for attentional selection and visual search. By the end of the introduction, I will again argue that our brains are well-equipped to deal with this challenge, and discuss how regularities in the real world can shape attentional templates and preparatory content-specific biases for efficient attentional selection and search.

An object rarely comes alone: the challenge of processing more than one object

It is hard to start a thesis on attention without mentioning the finite processing capacities and limited resources of our brains, in contrast to the seemingly infinite amount of sensory information that the external world seems to bombard us with. Think of a buzzing street in the city centre, a supermarket, our own living room, or a dense forest. Clearly, our typical environments contain a multitude of different objects, which may additionally undergo dynamic changes. But what exactly is limited? One clear limitation is that we can only fixate and look at a small portion of space at once. Visual acuity is only high in the fovea, which constitutes a small part of our retina, and declines sharply in the periphery. In order to process an object within our surroundings in detail, we thus have to move our eye towards it, and a series of quick eye movements is needed in order to explore a scene.

Another important limitation is related to representing and processing multiple objects present within our visual field at once. When we have to process only single, isolated objects, e.g., in a vision science lab, it is fairly easy to detect their presence and to then identify them. In the presence of other objects, however, our ability to detect a particular object decreases with the number of other (distractor) objects in the scene (Treisman & Gelade, 1980; Wolfe et al., 1989). Object detection and recognition, especially in the periphery, is worse when an object is surrounded by other objects (Carrasco et al., 1995; Lleras et al., 2022; Whitney & Levi, 2011; Wolfe, 2020). This is an indication for limited neural processing resources for which multiple objects have to compete, which can be seen at the level of their neural representations in visual cortex.

When one object falls into a neuron's classical receptive field (the part of the external visual environment it is responsive to), it will drive the neuron's response, increasing the neuron's firing rate if the neuron is tuned or selective to its features. An object's neural representation is however degraded by other nearby objects. With more than one object appearing within the classical receptive field, the neuron's firing rate will resemble a weighted average of the response to the individual objects (Chelazzi et al., 1993; Reynolds et al., 1999; Sato, 1989; Zoccolan et al., 2005, 2007). This response normalization can be explained by inhibitory connections between neighbouring neurons (Reynolds & Heeger, 2009). Normalization is not only found for the response of single neurons in monkey visual cortex, but also for fMRI response patterns in humans, reflecting the response of a large population of neurons (Baeck et al., 2013; MacEvoy & Epstein, 2009; Reddy et al., 2009). While this process avoids the potential saturation of neural responses, it can also render the firing rate of individual neurons ambiguous for later readout.

The strength of these competitive interactions increases along the visual processing hierarchy, as receptive fields in higher-level visual areas become larger, encompassing a larger portion of space and therefore more objects (Buffalo et al., 2010; Kastner et al., 2001). Within those higher-level visual areas, competition is not strongly determined by spatial proximity between objects, but rather their representational similarity and distance in feature-space, as nearby cortical patches have similar selectivity (Bao & Tsao, 2018; Cohen et al., 2017; Franconeri et al., 2013; Kliger & Yovel, 2020; Yeh et al., 2024).

Across neurons and visual regions, the individual weights given to objects in the normalization process can vary, ranging from a perfect average to a non-linear max-operation, in which highly selective neurons respond exclusively to their preferred stimuli, independent of the presence of other objects (Kliger & Yovel, 2020; Zoccolan et al., 2005). This could make an object's representation invariant to clutter (Rousselet et al., 2003, 2004), and such a max-response pattern is mostly found for categories with dedicated category-selective regions (e.g., for faces or bodies; Erez & Yovel, 2017; Kliger & Yovel, 2020; Reddy & Kanwisher, 2007). Generally, however, clutter remains a problem for the representation of most objects outside of those select few categories, or when multiple targets of the preferred category appear.

Attention biases and resolves competition

When processing cluttered scenes, we are, however, not passively observing. Our behavioural goals dynamically shape which aspects of the retinal input is currently relevant. As I am sitting at my desk writing, the coffee cup placed on it is more important to me than the pair of scissors next to it, and failing to register an approaching car or bike when crossing a road is very likely to have more severe consequences than missing the flashy billboard on the other side. Adaptive behaviour requires mechanisms to prioritise processing of relevant input over irrelevant one, by selectively attending to particular features, objects or locations. Independent of whether attentional selection is based on features, spatial location, or objects, the neural response to attended input is enhanced, with a more selective and synchronous response across neural populations, resulting in better detection and discrimination, at the expense of irrelevant input (Buschman & Kastner, 2015; Carrasco, 2011). The strongest attentional effects are therefore observed when multiple objects are presented simultaneously. Visual search, i.e., identifying and locating predefined targets, placed at an unknown location among other distractor objects, is likely the most common form of attentional selection we engage in, and its neural and behavioural correlates have therefore been extensively studied. How does making one object task-relevant shape the response to multiple objects in a display?

In an early study investigating the neural correlates of selective attention, Chelazzi et al. trained macaque monkeys to perform a search task and direct saccades to target objects in a multi-object search display, while recording the activity of neurons within object-selective inferotemporal cortex (IT) (Chelazzi et al., 1993, 1998). The initial neural response to the search display was the same regardless of which object was the target (see Figure 1). However, within about 200 ms, neurons responded as if only the relevant target was present. For instance, their firing rate decreased when a non-preferred object was the target, even in the presence of another preferred object that would increase its firing rate when shown in isolation.

How was competition resolved in favour of the target? Within the same study, the researchers also observed a putative attentional biasing signal. Prior to the search display, a picture cue indicated the relevant search target on the upcoming trial. In the delay between cue and search display (and also even prior to the cue display when the cue was predictable), the baseline firing rate of many neurons was increased (by about 30-40%) when their preferred stimulus was the target, in the absence of any visual stimulation (Figure 1C). This sustained preparatory activity was not found when the cue merely needed to be fixated rather than used in a search task, such that it likely reflected an internal top-down bias rather than a purely sensory response to the cue. This top-down bias may mediate competition in favour of the target in the subsequent search display. Accordingly, this feature-specific delay activity was interpreted as the neural signature of an attentional template, an internal representation of the relevant target object (Desimone & Duncan, 1995). Similar preparatory biases were later found not only for relevant features, but also when attention was directed towards a specific location (Kastner et al., 1999; Luck et al., 1997), indicating that the selective pre-activation of visual cortex may be a general principle to prioritise processing of relevant input (Battistoni et al., 2017; Nobre & Serences, 2018).

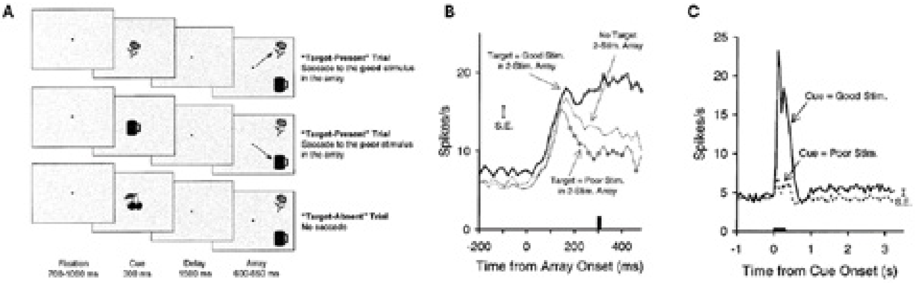

Figure 1: Neural responses of IT neurons during a search task. A Overview of the search task. Macaque monkeys were trained to make a saccade towards a cued target, if it was present. The target could either be a good (increasing the recorded neuron's firing rate) or a poor stimulus (not increasing the neuron's firing rate). **B** Average population response of neurons in inferotemporal cortex (IT) to the search display containing multiple objects. If neither of the two stimuli in the display is the target and none is attended (No Target), the response to the search display shows response normalisation. If the target is either the good or poor stimulus, the neural response resembles the response to this target object alone, even in the presence of the other object. **C** Average population response to the cue and during the delay period, showing an increase in the baseline firing rate when the cued target is the good stimulus. All panels are taken from Chelazzi et al. (1998).

These findings laid the basis for the biased competition theory of attention (Chelazzi et al., 2011; Desimone & Duncan, 1995), which includes two main tenets. First, objects falling into a neuron's receptive field will compete for driving the cell's response and therefore for neural representation. Second, this competition can be biased in favour of one of the competing objects, either through bottom-up factors (e.g., when one stimulus has a higher contrast or is brighter), or top-down factors (i.e., when one is more behaviourally relevant). These claims are also consistent with other influential theories of attention, and the biased competition theory can be seen as a particular instance of the normalisation model of attention (Reynolds & Heeger, 2009). This states that the excitatory drive of a stimulus on a neural population is normalised by the cumulative suppressive influence pooled across all nearby objects. Prior to this normalisation, attention acts by scaling the excitatory drive of the attended object and thereby increasing its gain. The biased competition theory, however, proposes a clear neural implementation of the attentional biasing signal reflected in preparatory activation of target-selective neurons.

Stages and neural mechanisms of visual search

Understanding the basic mechanisms of attentional prioritisation has been one of the main goals of cognitive neuroscience in the past decades, and using simple search tasks similar to the one just described, there is now growing consensus about general stages of search and factors determining the efficiency of attentional selection (Eimer, 2014; Wolfe, 2020, 2021).

Search begins by deciding what to look for, i.e., by creating an internal representation of the target, termed attentional template or also search template. Within lab-based search tasks, the target is typically explicitly provided, as picture, word, or symbolic cue defining the target for the next trial or block of trials. As emphasised by the name, the attentional template is assumed to bear some similarity to the visual representation of the target. The amount of prior information we have about visual features of the target are a crucial factor determining search efficiency and how attention is allocated (Bravo, 2009; Malcolm & Henderson, 2009; Maxfield & Zelinsky, 2012; Schmidt & Zelinsky, 2017; Vickery et al., 2005; Wolfe et al., 2004). This target representation is most commonly thought to be held in visual working memory and instantiated by top-down (pre)activation of target-selective neurons in visual cortex (Battistoni et al., 2017; Desimone & Duncan, 1995). Throughout the search process, objects are prioritised for processing based on their match with this template.

Once the template is instantiated, we do not blindly search for template-matching objects at every location within our visual field in a serial manner. Instead, attention is guided towards likely target candidates and their locations, while suppressing unlikely ones. Evidence for the presence of simple target features (e.g., red, vertical, or rightward moving items) can be accumulated within a single glance (Treisman & Gelade, 1980; Wolfe, 2021). The efficiency of guidance, measured by the proportional increase in reaction time with the number or distractors (the search slope) depends on how well the target can be separated from distractors along those feature dimensions. Search is easiest if the target is unique and distinct from distractors (i.e., target-distractor similarity is low) and if distractors themselves are homogeneous (distractor-distractor similarity is high) (J. Duncan & Humphreys, 1989), making the target “pop out” of the search scene. If feature-based guidance is inefficient, however, almost every item in the display needs to be foveated and attended sequentially to distinguish it from distractors. At the neural level, attending to a particular feature, i.e., deploying feature-based attention, increases the response gain of feature-selective neurons, across the entire visual field (Bichot et al., 2005; Maunsell & Treue, 2006), thereby highlighting potential targets independent of their location. Feature-guidance might be restricted to relatively simple features such as colour, motion, or orientation (Wolfe & Horowitz, 2004, 2017), for which there is also evidence for global attentional gain modulation (Bichot et al., 2005, 2019; T. Liu et al., 2007; Saenz et al., 2002; Serences & Boynton, 2007; Treue & Martínez Trujillo, 1999). Recent studies have, however, also found evidence for spatially global gain enhancement for some of the most relevant high-level object categories: human bodies and faces (Störmer et al., 2019; Thorat & Peelen, 2022).

Following the detection of likely target candidates, spatial attention is shifted towards their location for further, focused analysis, allowing to identify the target. This spatial selection is reflected in an enhanced neural response for the selected object across the ventral visual cortex within about 200 ms (Chelazzi et al., 1993; Reynolds et al., 1999), independently and additive to the previous feature-based gain-enhancement (Treue & Martínez Trujillo, 1999). This spatial selection is also indicated by enhanced perceptual sensitivity at the attended location (Downing, 2000; Reeder & Peelen, 2013), neural markers of spatial selection such as the N2pc event-related potential (Luck, 2012; Luck et al., 2000; Stoletniy et al., 2022), and overt eye movements. When the selected object is identified, the searcher will need to make a decision whether it is indeed the target. Search is deemed successful if it matches the memorised target representation. If it does not, attention may shift to the next potential candidate, or search is terminated if the searcher determines that the target is unlikely to be present.

What I have described so far has been derived mostly from search experiments within relatively controlled and artificial environments, looking for, e.g., oriented and coloured bars, letters or isolated objects that are arbitrarily arranged in the display. More naturalistic searches differ from those lab-based ones in some important aspects (Peelen & Kastner, 2014; Wolfe, Alvarez, et al., 2011). For instance, we have not yet discussed how spatial information provided by a naturalistic scene contributes to guidance. Despite these differences, the very general stages and principles described here are also found when searching through real-world scene photographs rather than arbitrary search displays. Prior to the onset of the search scene, preparatory activity in object-selective cortex reflects categorically-defined search targets such as people, cars, or other objects (Gayet & Peelen, 2022; Peelen & Kastner, 2011; Soon et al., 2013). Soon after the search display is shown, the representation of the target category is enhanced (Kaiser et al., 2016; Peelen et al., 2009; Störmer et al., 2019; Thorat & Peelen, 2022), before spatial attention is drawn towards the location of target objects, allowing for identification (Battistoni et al., 2020).

Besides incorporating scene-guidance, the latest models of search, such as the current version of the Guided Search Model (GS 6.0), also propose a distinction between target- and guiding templates (Wolfe, 2021). Until recently, it was generally assumed that a single unitary target representation, encoded in the attentional template, supports both attentional guidance and identification (Hout & Goldinger, 2014). Contrary to this view, and as we will discuss in more detail throughout this introduction, target and guiding-representations could in fact be dissociable, especially in search for more complex detailed objects and extended search with many targets. One reasons for this is that the memory representation of the target used for identification (measured in unspeeded decision tasks) can be more precise and/or slightly distinct from the features used for guidance, as revealed by eye-movements or EEG markers (S. I. Becker et al., 2023; X. Yu, Hanks, et al., 2022; X. Yu, Johal, et al., 2022). This is related to the fact that not all features useful for discrimination can be used to efficiently guide attention (Wolfe & Horowitz, 2004, 2017), for instance because they cannot be efficiently processed within the periphery (Lleras et al., 2022) and/or do not have a retinotopic organization that may be necessary for the global deployment of feature-based attention (Maunsell & Treue, 2006). Instead of a perfect, veridical target representation, the target-representation used for guidance may therefore simply be "good-enough", as errors in guidance typically do not come at a huge cost, while the representation used for target identification should be more precise and veridical (X. Yu, Zhou, et al., 2023). Another reason to dissociate both is the fact that the number of targets we can memorise and effectively match to visual input by far exceeds the capacity of working memory and guiding representations (Shang et al., 2024; Williams et al., 2024; Wolfe, 2012, 2020). How target information and these potentially distinct templates are encoded in the brain therefore becomes a highly relevant question.

Preparatory activity in search tasks and beyond

Following the initial discovery of preparatory activity in macaque visual cortex (Chelazzi et al., 1993), a range of neuroimaging studies in humans have found similar pre-activation of target-representations in visual cortex across a variety of search tasks and in multiple areas within the visual hierarchy (Battistoni et al., 2017). Early fMRI studies showed overall increases in baseline activity of different feature-selective regions, depending on whether the target was defined by either colour or motion (Chawla et al., 1999; Giesbrecht et al., 2006). Multivariate decoding methods allowed for a finer-grained analysis of the content of preparatory activity within a given region, providing evidence for preparatory representations of simple features in early visual cortex (Chen et al., 2024), as well as for letters (Stokes, Thompson, Nobre, et al., 2009), and complex real-world categories and objects in object-selective occipito-temporal regions (Gayet & Peelen, 2022; Peelen & Kastner, 2011; Soon et al., 2013; Zhou & Geng, 2025).

In order to provide clear evidence for an attentional biasing signal in the absence of visual stimulation, any preparatory activity should be clearly separated from the visual response to the target, which is especially challenging in fMRI given the overall sluggish hemodynamic response underlying this signal. This can, however, be done by either introducing separate catch trials, in which participants are cued and anticipate to search but the search scene is then withheld, or by including long delay periods allowing to clearly dissociate the BOLD response during cue, delay period, and search scene. The preparatory templates found in these studies were sensory-like, resembling visually driven responses to the targets. When classifiers were trained to distinguish the current target from preparatory activity, those classifiers could generalise and successfully distinguish between visually evoked responses by the two targets. Importantly, some studies found that the strength of those preparatory activity patterns correlated with behaviour across participants (Peelen & Kastner, 2011; Soon et al., 2013; Stokes, Thompson, Nobre, et al., 2009). Transcranial magnetic stimulation (TMS) over category-selective regions following a cue also disrupted later perceptual performance for targets of that category, providing causal evidence for the functional role of preparatory biases in visual cortex (Gandolfo & Downing, 2019; Reeder et al., 2015).

Attentional biases observed in sensory regions are thought to reflect top-down modulation by frontal control regions (Desimone & Duncan, 1995; Martinez-Trujillo, 2022). The inferior frontal junction (IFJ, called ventral prearcuate region (VPA) in nonhuman primates) contains feature-selective neurons that show persistent attentional effects throughout the search trial, arising earlier than in visual cortex (Bichot et al., 2015). In addition, inactivation of VPA neurons abolishes feature-based attentional gain enhancement in visual cortex, making it a likely source region of content-based attentional templates (Bichot et al., 2015, 2019). The frontal eye field (FEF), involved in the control of eye movements, also has retinotopic projections to visual cortex (Gregoriou et al., 2012; Moore & Armstrong, 2003; Ruff et al., 2006) and has been linked to spatial biasing signals also in the absence of overt eye movements. Besides frontal regions, also areas within the posterior parietal cortex, such as the intraparietal sulcus (IPS) can provide feedback to the ventral visual stream (Y. Xu, 2023). In addition, feedback projections originating in the hippocampus may also modulate visual cortex activity during search preparation, when attentional prioritization is based on memory and learnt spatial associations (Aly & Turk-Browne, 2017; Favila & Aly, 2023; Günseli & Aly, 2020; Stokes et al., 2012).

As is clear from the review above, preparatory activity in visual cortex reflects a crucial cornerstone in current theories of selective attention (Buschman & Kastner, 2015; Carrasco, 2011; Desimone, 1995). However, despite growing experimental evidence for content-specific preparatory biases in visual search tasks, there are still open questions regarding their interpretation and functional relevance as an internally driven attentional biasing signal, particularly for real-world search.

First, an important question is whether attentional templates are always encoded through sustained firing of sensory neurons. Such a neural code may for instance be disrupted by concurrent sensory input. Some studies have provided evidence that top-down signals, especially within early visual cortex, are abolished by distractors in the delay period and are only persistent within parietal or frontal areas (Bettencourt & Xu, 2015; Miller et al., 1996; Y. Xu, 2020). Others have, however, shown that, while this may come at some cost in terms of representational fidelity compared to no distraction, both can simultaneously co-exist even within sensory regions, and the effect of disruption is reduced when moving along the visual hierarchy (Iamshchinina, Christophel, et al., 2021; Kiyonaga et al., 2017; Rademaker et al., 2019; Rademaker & Serences, 2024). This is relevant as many attentional templates for search in real-world scenes may be for whole objects or object categories and are encoded within those higher-level visual areas. Time-resolved analyses have also shown that top-down signals are briefly disrupted by concurrent visual input, but can be re-instantiated quickly (Hallenbeck et al., 2021; Kerkoerle et al., 2017), suggesting that top-down activity in sensory cortices can be present also during concurrent processing of a scene, as would be the case in real-world search. However, it is also possible that attentional templates are not always persistently encoded in sustained activity of target-selective neurons across longer durations. For instance, an item that we will later have to search for, but is not immediately relevant, can be encoded in visual cortex, but its neural code may differ from when it becomes relevant as a search template (de Vries et al., 2017; Loon et al., 2018; Q. Yu et al., 2020). Similarly, a recent study (Chen et al., 2024) suggested that attentional templates may be encoded in an ‘activity-silent’ latent state, reflected in synaptic plasticity rather than active firing, which could reduce both metabolic cost and distractor interference (Mongillo et al., 2008; Stokes, 2015). In this study, an active and decodable sensory-like template, however, emerged once visual input was presented to probe these latent representations. It is important to note that the prior absence of above-chance decoding by itself does not necessary imply a different type of code, but could also still reflect a weak, but sustained signal (Barbosa et al., 2021). These varying codes may substantially mitigate the potential impact of concurrent sensory input, especially when it is unlikely to already contain the target. Overall, while a sensory-like preparatory template may become especially useful once the template needs to interact with bottom-up visual input, information about relevant search targets appears to be encoded in visual cortex, prior to search, also in the presence of concurrent sensory input.

Second, the functional role of preparatory activity, and how it is linked to the different stages of search, is not yet clear. More specifically, does preparatory activity support attentional guidance and/or target identification? First, there is a strong theoretical link between preparatory biases and early attentional processes such as guidance. Preparatory activity is thought to resolve competition between objects in favour of the target, thereby providing efficient attentional guidance, prior to any potential identification processes. Pre-activation of visual neurons should directly enhance the subsequent attentional gain, linking preparatory activity to effective feature-based attention. There is indeed some evidence showing that the same neurons active prior to search onset also show increased attentional gain (Chelazzi et al., 1993, 1998). On the other hand, this link between preparatory activity and attentional guidance is not always clearcut. There is some evidence that the neurons showing preparatory activity during a delay period and those showing feature-based attentional gain can be distinct from each other (Mendoza-Halliday et al., 2024). Could preparatory activity then reflect a template for target identification only? Target-selective activation during a delay period can indeed also be found in working memory and decision-making tasks, in which the target needs to be memorised and identified, but there is no simultaneous competition from distractors or need for guidance (Christophel et al., 2017; Dodwell & Eimer, 2024; Miller et al., 1993; Myers et al., 2015; Stokes, 2011). Overall, when it comes to the potential distinction between target and guiding templates, it is still an open question whether they can be distinguished at the neural level, and whether preparatory activity reflects the neural basis for either one. One important issue in distinguishing between both is that in prior studies investigating preparatory activity, the features relevant for guidance have always been conflated with those relevant to identify the target, i.e., typically only veridical features of the cued target could be used for both identification and guidance, and those were the features reflected in preparatory activity. Yet, as already mentioned, attentional guidance may rely on a distorted representation of the target, or even entirely different features (X. Yu, Zhou, et al., 2023). To provide clear evidence for or against the functional role of preparatory activity in attentional guidance, this attentional biasing signal needs to be isolated from other cue- and target-related processes (Moerel et al., 2022). In Chapter 3, we will dissociate target and guiding features to investigate the role of preparatory activity as a guiding template.

Top-down activation of visual cortex during working memory, mental imagery, and expectation

Top-down activation of visual representations in the absence of visual input has not been found exclusively in search tasks, but has also been found in visual working memory tasks (Christophel et al., 2017; Gayet et al., 2018; Harrison & Tong, 2009; Iamshchinina, Christophel, et al., 2021; Miller et al., 1993; Serences et al., 2009), during mental imagery (Albers et al., 2013; Cichy et al., 2012; Dijkstra et al., 2017, 2019; Naselaris et al., 2015; Stokes, Thompson, Cusack, et al., 2009; Wadia et al., 2024), and has also been linked to expectations (Kok et al., 2014, 2017), with overall similar facilitatory effects on perception (Farah, 1989; Gayet et al., 2013; Pashler & Shiu, 1999; Stein & Peelen, 2015). Is preparatory activity in visual search tasks therefore "just" working memory or imagery (see also Battistoni et al., 2017; Stokes, 2011)? What are the interconnections between those processes, and how may they be distinct?

Upon the discovery of preparatory activity, the similarity to delay-period activity in working memory tasks, in which a memorised item needed to be compared with a single probe item (Miller et al., 1993), was immediately noticed, prompting the researchers to claim "visual search simply appears to be a variant of a working memory task, in which the distractors are distributed in space rather than in time." (Desimone & Duncan, 1995, p. 205). As target features are not directly present when we begin to search, the search process clearly involves some memory for the target. There is indeed a tight link between visual working memory and attentional selection, as objects matching working memory representations can automatically capture attention (Downing, 2000; Kumar et al., 2009; Olivers, 2009; Olivers et al., 2006, 2011;

Soto et al., 2005, 2008) and attract eye movements (Bahle et al., 2018; Beck et al., 2012; van der Stigchel & Hollingworth, 2018), are prioritised for conscious access (Gayet et al., 2013; Moorselaar et al., 2018), and show an enhanced neural response compared to non-matching objects (Bahmani et al., 2018; Gayet et al., 2017). However, not every working memory representation acts as a search template. Working memory content may be encoded using a variety of coding schemes, including, but not limited to, sustained firing activity within visual cortex (Chota & van der Stigchel, 2021; Christophel et al., 2017; Sreenivasan & D'Esposito, 2019). For instance, working memory representations may be found only in parietal or frontal regions, or also in previously mentioned 'latent' activity-silent states, mitigating the potential impact of concurrent sensory input in visual cortex (Lorenc et al., 2018; Stokes, 2015). In line with this, not all items within working memory capture attention (or at least do so to a lesser degree), when there is no need to search for them (Carlisle & Woodman, 2011; Downing & Dodds, 2004; Woodman & Luck, 2007). Pure memory and search tasks also differ in their respective task demands. While the goal in the former is to maintain a veridical memory trace, which may need to be shielded from external interference, attentional templates should reflect an optimised target representation to distinguish targets from distractors that can be continuously matched to visual input. These differences likely influence the specific content and code of internal representations and their interactions with visual input (Olivers et al., 2011; Stokes, 2011). In addition, visual search, especially for familiar targets or in familiar environments, would also recruit long-term memory (Aly & Turk-Browne, 2017; Carlisle et al., 2011; Gunseli et al., 2014; Hutchinson & Turk-Browne, 2012).

The sensory-like nature of attentional templates provides an intuitive link between preparatory attentional templates and mental imagery (Pashler & Shiu, 1999). Engaging in visual imagery can be an adaptive and useful strategy in working memory or search tasks, especially when the target representation needs to be highly detailed to successfully complete the task. However, imagery is often effortful and we do not typically experience a vivid visual image in front of our minds' eyes when searching for objects in scenes. Creating a vivid mental image may rely on top-down activation of especially early visual cortex (Dijkstra et al., 2017), but such an 'imagery-like' template has been found to even hinder search for categorically defined targets that vary in their visual features (Bravo & Farid, 2014; Peelen & Kastner, 2011). Another line of evidence for the dissociation between explicit and voluntary imagery and attentional templates comes from the study of aphantasics, people who do not experience conscious voluntary imagery (Zeman, 2024). Similar to normal imagers, aphantasics' attention is still captured by matching colour patches when looking for familiar fruits, suggesting that they created preparatory attentional templates including colour information (Cabbai et al., 2023). In addition, there is recent evidence for top-down activation of visual representations in aphantasics or otherwise bad imagers outside of imagery tasks, which may potentially be less sensory-like, but again calls for a dissociation between imagery and working memory or attention (Cabbai et al., 2024; Chang et al., 2025; Weber et al., 2024).

A final important distinction between imagery and preparatory activity is that, in contrast to attentional templates, imagined objects can be entirely unrelated to the external world. Rather than being focused on the external world, attention is turned inwards towards these internally-generated representations (Chun et al., 2011; van Ede & Nobre, 2023). In that case, imagery content and perceived objects may co-exist in visual cortex without the need to interact. Given this co-existence and the representational overlap between imagery and perception, a so-far little explored consequence is that both could also compete for representation in visual cortex, as suggested by a recent computational model (Sulfaro et al., 2023). With concurrent visual input, imagery may therefore be less strongly represented across visual cortex, especially less likely to reach and activate lower-level visual areas (see Figure 1), consistent with the previously discussed findings showing visual stimulation can degrade internal representations in especially early visual cortex (Iamshchinina, Christophel, et al., 2021; Lorenc et al., 2021; Rademaker et al., 2019; Rademaker & Serences, 2024). The other side of this coin is that attending to internal representations may also interfere with concurrent perceptual processing. Importantly, this competition should also depend on representational similarity and neural overlap, i.e., the degree to which imagined and perceived objects are drawing on the same limited neural resources, with more similar representations resulting in stronger mutual interference. Is it possible that we see an object less clearly when we are concurrently imagining a similar object? In Chapter 4, we will test this hypothesis.

Finally, the mere *expectation* of a particular stimulus or category also induces increases in baseline activity within visual regions (Esterman & Yantis, 2010; Puri et al., 2009), encoding content-specific representations (Hetenyi et al., 2024; Kok et al., 2014, 2017; Noah et al., 2020). Similar to the biasing effects of attentional templates, expected or probable visual input may also be prioritised for conscious access (Stein & Peelen, 2015; L. Xu et al., 2024). It is, however, possible to experimentally dissociate attention and expectation by independently manipulating the probability of a stimulus to appear (expectation) and its relevance to the task (attention) (C. Summerfield & de Lange, 2014), and studies of preparatory attention in search have therefore equalised the overall prevalence of target categories and vice versa for studies of expectations. Despite this distinction, attention and expectations may of course interact, for instance the attentional template may be shaped by contextual expectations, as we will discuss in the following section.

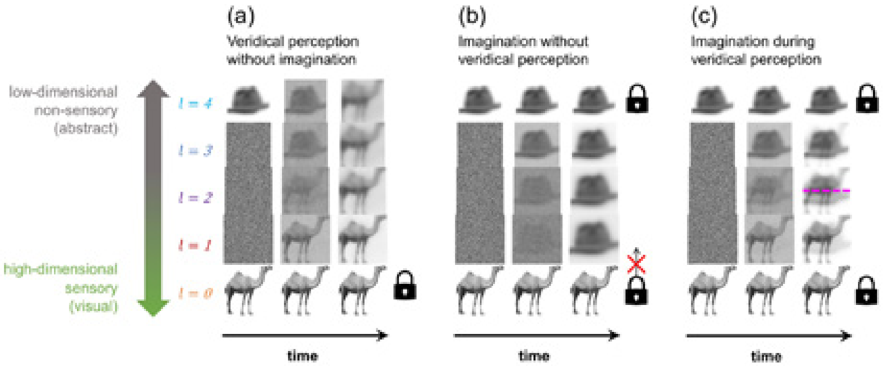

Figure 2: Model for interference between imagery and perception. Simulated neural representations of visual and imagined input across time and the visual hierarchy. Layers L0 – L4 represent different levels of the visual hierarchy. **A** For perception without concurrent imagination, bottom-up input is processed at all levels of the visual hierarchy. **B** Imagery without concurrent perception leads to representation of the imagined content across all levels of the visual hierarchy. **C** For imagery with concurrent perception, imagery and perception compete, resulting in perceptual input not being processed in higher-level layers, while imagined input does not reach lower-level areas. Figure taken from Sulfaro et al. (2023).

Overall, top-down activation of stimulus-specific representations in visual cortex in the absence of visual input is commonly observed across a variety of tasks, especially when precise visual features of a stimulus need to be encoded and maintained. However, while different tasks can lead to superficially similar delay-activity, the evoked internal representations are not necessarily identical, and could differ in terms of, e.g., their interactions with external input, their specific neural code, and also their source regions.

Using environmental regularities for efficient search

One of the main appeals of studying visual search in the lab is that search is ubiquitous in real-world behaviour. We are almost constantly engaged in search, looking for our glasses on the nightstand, for a ripe apple in the supermarket display, for a friend at a concert, and so on. Natural environments are highly cluttered and complex, yielding rather pessimistic predictions about the efficiency of real-world search. In addition, the specific features of the target are often uncertain and ill-defined. There is substantial variation between objects of the same category, and even the very same object may look drastically different at varying distances, viewpoints and illumination conditions. Yet, barring the occasional frustration of not seeing an object right in front of our nose, we typically perform all these searches efficiently, and seemingly without much effort (F. F. Li et al., 2002; Potter, 1975; Thorpe et al., 1996; Wolfe, Alvarez, et al., 2011), demonstrating that our brains are finely tuned to process naturalistic scenes and search for real-world objects within them. Despite their apparent complexity, an important feature of real-world environments is that they are highly structured and predictable, offering a wealth of spatial and object-based regularities, and we have a lifetime of experience navigating them (Kaiser et al., 2019; Peelen et al., 2023; Peelen & Kastner, 2014; Võ, 2021). When looking for an apple in my friend's kitchen, I can capitalise on the fact that they are typically red, about as large as my fist and are commonly found in fruit bowls placed on kitchen counters, perhaps lying right next to some yellow bananas. If I even happen to remember the shape and location of the given fruit bowl from a previous visit, I will have immediately found the apple. This should be a major contributing factor to the efficiency of real-world search. Indeed, search performance is typically worse when such regularities are broken (Biederman et al., 1973, 1982; Castelhano & Heaven, 2011; Eckstein et al., 2006, 2017; Spotorno et al., 2014). In the following sections, I will review work investigating how searching structured environments may shape attentional guidance and the target template.

Adapting the template to expected distractors

When searching through a naturalistic scene, we often have some knowledge not only about the target, but also about the likelihood of encountering different non-target objects. This is relevant as the efficiency of search is not based on features of the target alone, but based on features that distinguish the target from the distractors (J. Duncan & Humphreys, 1989). In principle, distractor expectations could be used in different ways for efficient search (van Moorselaar & Slagter, 2020): either directly suppressing distractors in a proactive (i.e., prior to display onset; Arita et al., 2012) or reactive manner (Gaspelin & Luck, 2018; Moher & Egeth, 2012), or alternatively re-shaping the attentional template in a way that reduces overlap with distractor features. Here, we will focus on the latter.

An effective attentional template should highlight specifically those features which reliably characterise the target and set it apart from expected distractors. When we are looking for objects of familiar categories, our search templates are likely already optimised to include diagnostic information distinguishing it from most commonly encountered distractors following extensive experience (Sigala & Logothetis, 2002).

For instance, when searching for categorically defined targets such as people or cars, attention is captured by category-diagnostic midlevel features (e.g., the shape of an arm or leg) but not the colour or texture of individual items, which are less characteristic of the respective categories across a broad range of exemplars (Reeder & Peelen, 2013).

There is also evidence for such adjustments over much shorter timeframes, in the range of a few trials. When looking for odd-one out targets among distractors, participants quickly adapt to the specific shape of the feature-distribution distractors are sampled from, with increasing reaction times when the distractors on test trials do not match this distribution (Chetverikov et al., 2016, 2017, 2019). When searching for more complex multidimensional targets, less variable or more diagnostic feature dimensions (e.g., because target colour, but not orientation is distinct from distractors) appear to be emphasised in the template. This leads to worse search performance on trials in which distractors share the expected diagnostic dimension (Boettcher et al., 2020; J. Lee & Geng, 2020; Witkowski & Geng, 2019).

Is the attentional template for real-world object categories also this flexible? Object representations in ventral visual cortex have been shown to be relatively independent of specific task context (Bracci et al., 2017; McKee et al., 2014). Their representations may therefore be more integrated and less malleable. However, other studies have also provided evidence for more task-dependent representations in those areas (Harel et al., 2014; Henderson et al., 2023; Y. Xu, 2023). There is also some evidence that when searching for such real-world objects, the degree of detail included in the target template can vary when anticipating a more difficult search and depending on recent distractor experience (Addleman et al., 2024; Bravo & Farid, 2012; Lau et al., 2021; Schmidt & Zelinsky, 2017). This indicates that searchers rely on different target information depending on the expected distractor context and search difficulty, even during real-world search.

While there is growing evidence that the template can be shaped by anticipated distractors and search difficulty, it is useful to distinguish between different potential mechanisms behind attentional flexibility and learning of environmental regularities. Imagine looking for a colleague, who is almost always wearing a baseball cap, such that it becomes a diagnostic feature and included in our attentional template. This allows for efficient search at an academic conference, but we would be less successful looking for the same colleague at a hip-hop concert the week after with the same template, given that wearing a cap is not a unique feature of our colleague among the concert-goers. One possibility is that a change in the template arises as a simple consequence of passive adaptation, reducing the overall activity of any neurons responsive to caps when they are abundant (Turatto et al., 2018). Another possibility is that frequent attentional capture by distractor-caps will trigger reactive control processes and a strategic adjustment of the template. In both cases, the neural gain in response to caps would be effectively reduced, compared to other features included in the template (e.g., the colour of the hoodie we saw him wearing today), resulting in an adaptive change in the template. It is becoming increasingly clear that the immediate past and recent search history strongly impacts attentional deployment (Anderson et al., 2021; Awh et al., 2012), and this can lead to adaptive changes in the attentional template within stable environments (Kristjánsson, 2022). The majority of the studies reviewed here and in the following sections investigated regularities that remained fixed over blocks of trials, the whole experiment, or based on lifelong learning. The current example, however, also shows that diagnostic features of the target change in different contexts and it would thus be useful to learn regularities in a contextualised manner, meaning that learning of diagnostic features is tied to specific contexts, and flexibly used and retrieved directly based on context-cues. How flexible and truly context-dependent can the attentional template be?

Evidence for such flexible, context-driven adjustments in the attentional template, and specifically attentional guidance, is relatively scarce so far. In Chapter 2 and 3, we will investigate how the attentional template and preparatory activity can be flexibly adjusted based on trial-by-trial context cues.

Optimal and "good-enough" templates for guidance and decision-making

Accounting for the feature distribution of distractors could be especially useful to support fine-grained feature discriminations and difficult searches, e.g., when looking for the ripest and reddest apple in the supermarket display. So far, we have assumed that the guiding template should be a somewhat coarser, perhaps slightly distorted (in order to emphasise diagnostic target features), but overall veridical representation of the actual target features. This would, however, not be very helpful for this task since a template for red, round objects would select almost all objects equally. As we can see from the colour tuning curves in Figure 3A, the signal-to-noise ratio (SNR) between neurons coding for the actual target colour-value (1) and distractor apples is very low, but can be increased when neurons tuned to an even redder colour value (2), redder than any apple in the display is considered. Thus, the sensory gain of those neurons should be enhanced and the attentional template shifted away from distractors (Geng & Witkowski, 2019; Navalpakkam & Itti, 2007). To test this, researchers have asked participants to search for predefined targets among distractors that are consistently shifted along one axis in feature space across trials. Optimal tuning, or off-target enhancement, has been found for orientation (Navalpakkam & Itti, 2007; Scolari et al., 2012; Scolari & Serences, 2009, 2010), colour (Geng, Quattro, et al., 2017; Kerzel, 2020; Schulz et al., 2024; X. Yu, Hanks, et al., 2022; X. Yu, Johal, et al., 2022; X. Yu & Geng, 2019), speed and direction of motion (Boettcher & Nobre, 2024), and even emotional facial expressions (Won et al., 2020). A related idea is that the template may not encode a specific off-target feature value, but rather is tuned to the overall relation between targets and distractors (i.e., redder) (S. I. Becker, 2010b, 2013; S. I. Becker et al., 2013). These shifts in the template appear to occur only when it is it is useful for search rather than reflecting automatic perceptual contrast effects (X. Yu, Rahim, et al., 2023). In this study, participants searched for predefined targets differing in colour and orientation. For the same distractor colour values, the colour encoded in the target template was shifted away from distractors mostly when the participants could not rely on orientation information to easily dissociate targets from distractors.

Importantly, this series of studies also showed a clear distinction between the underlying memory representations of the target and features used for guidance. Using different approaches to measure the internal target representation yields interesting diverging results. Measures of neural gain after extensive training typically show a gain-enhancement for neurons tuned quite far away from the current target features (for instance up to 36° in orientation), reflecting those neurons with the most discriminative response between targets and distractors (Jazayeri & Movshon, 2006; Pouget & Bavelier, 2007; Purushothaman & Bradley, 2005; Scolari et al., 2012; Scolari & Serences, 2010). Measuring attentional guidance through overt eye movements or the N2pc also yields a relatively broad template (S. I. Becker et al., 2023; X. Yu, Hanks, et al., 2022; X. Yu & Geng, 2019), while assessing the participant's memory for the target, e.g., through an unspeeded choice task or asking them to reproduce the exact target colour on a colour wheel will show a smaller, but still significant shift (Chapman et al., 2023; X. Yu & Geng, 2019). This shows that the guiding template includes features that are 'good-enough' for fast and efficient attentional guidance, rather than accurately reflecting target values, while the target representation remains at least close to target values (X. Yu, Zhou, et al., 2023).

Attentional guidance by the scene, spatial regularities and object co-occurrence

Target features are only one source of spatial guidance in real-world scenes (Castelhano & Krzyś, 2020). Within a single glance of a scene, its 'gist', including coarse spatial layout, i.e., whether a scene is open or closed, and basic category, e.g., forest or cityscape can be extracted (Fei-Fei et al., 2007; Greene, Michelle & Oliva, Aude, 2009; Henriksson et al., 2019). This information is based on global features and summary statistics (Torralba et al., 2006; Wolfe, Võ, et al., 2011), and can be used to constrain search to likely locations and surfaces (Pereira & Castelhano, 2019). Providing a preview of the scene prior to the scene helps to extract this layout and leads to efficient guidance (Castelhano & Heaven, 2010; Castelhano & Henderson, 2007; Võ & Henderson, 2010), and participants show considerable consistency in the locations of the first fixation in a scene when searching for a particular object, even when the target is actually absent from the scene (Eckstein et al., 2006; Ehinger et al., 2009; Torralba et al., 2006). Finally, search for, e.g., an airplane is more efficient when it appears in the upper rather than the lower visual field (Neider & Zelinsky, 2006; Spotorno et al., 2014).

Learning of spatial regularities has also been found over much shorter timescales and simpler scenes. Crucially, for real-world search it would again be useful if this learning is contextualised, i.e., depending on the specific combination of target identity, location, and scene context. There is so far mostly partial evidence for this, indicating the combined learning of two of those three elements respectively. One prominent example of learning associations of scene context and location is contextual cueing, in which participants are typically asked to look for a T-shaped target placed among L-shaped distractors. Unbeknownst to them, some of the search displays are repeatedly encountered while others are novel, and the spatial configuration of distractors in the repeated scenes are thus predictive of target location. Despite participants typically reporting to be unaware of these repetitions, search in repeated displays shows consistent RT-benefits over novel displays, at least partially reflecting more efficient attentional guidance to the target location (Bouwkamp et al., 2024; Brady & Chun, 2007; Chun & Jiang, 1998; Sisk et al., 2019). Similarly, the location of targets superimposed on real-world scenes can be memorised and explicitly retrieved, even for up to hundreds of scenes, and used for attentional guidance (M. W. Becker & Rasmussen, 2008; Brockmole & Henderson, 2006; Stokes et al., 2012; J. J. Summerfield et al., 2006), indicating that they are powerful contextual cues.

In addition, many studies have shown that repeating the target location across trials makes search more efficient and influences attentional guidance (Geng & Behrmann, 2005; Golan & Lamy, 2024; Theeuwes et al., 2022), also modulating the excitability of visual cortex in a retinotopic manner prior to display onset (D. H. Duncan et al., 2023; Ferrante et al., 2023). Interestingly also, it has recently been shown that these probabilities can be tied to the identity of specific targets (Zhang & Carlisle, 2023). With extensive experience, these location probabilities may become firmly etched into the object representation and attentional template for real-world object categories. In line with this, detection of isolated objects with strong positional regularities, e.g., ceiling lamps or carpets are better detected when their location is consistent with their typical visual field location (Kaiser et al., 2019; Kaiser & Cichy, 2018), suggesting spatial tuning for those object-selective neurons.

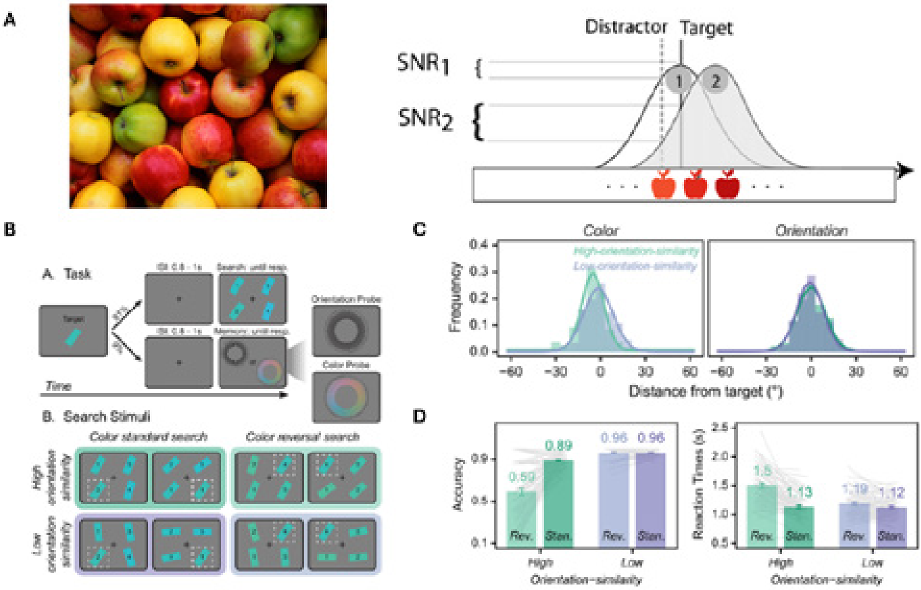

Figure 3: Optimal and off-veridical templates. A When looking for the reddest apple in this scene, neurons tuned to the actual target value (1) show similar activation for targets and distractors and SNR is thus low, but can be increased when neurons are tuned away from distractors, towards a more extreme, off-target value (2). The right panel is adapted from Navalpakkam & Itti (2007). **B** Overview of the search task in Yu, Rahim et al. (2023). Participants searched for targets defined by colour and orientation. The target is unique in colour and orientation on each trial, but only target colour is predictably shifted along one direction across trials. Orientation can be easily used as diagnostic dimension for the low-orientation similarity group, but not the high orientation similarity group. The target template is assessed by probe trials in which either colour or orientation of the target needs to be reported and performance on rare colour reversal trials **C** Distributions of responses for the colour and orientation probe trials, showing that memory for the target colour is shifted away from distractors especially in the high-orientation similarity group, and similar representations of the target orientations across groups. **D** Performance in the high-similarity group was impacted by colour-reversal trials, in which distractor-colour was close to the off-veridical target value. B,C and D are taken from Yu, Rahim et al. (2023).

Until now we have mostly considered the global or absolute position of objects within a scene or on a computer display. Spatial predictions can also be derived from the co-occurrence between objects and their relative positions. Going back to our example of looking for apples, we know that they typically occur on top of kitchen counters. The kitchen counter reflects a so-called "anchor object", a large and salient object reliably associated with other local objects (Boettcher et al., 2018). Such object associations are also used in attentional guidance, especially for smaller less salient targets (Draschkow & Võ, 2017; Helbing et al., 2022; Koehler & Eckstein, 2017a, 2017b; Mack & Eckstein, 2011; Zhou & Geng, 2024). Consistent object associations reshape the neural representation of individual objects towards a more integrated representation (Kaiser & Peelen, 2018; Schapiro et al., 2012), which could support the retrieval of associated objects during search (Kallmayer et al., 2024). Guidance by anchor objects also demonstrates that the guiding template may in some cases not reflect the search target itself at all. In Chapter 3, we will investigate the content of preparatory activity in context-guided search to test its role as a guiding template.

Adjusting preparatory activity based on contextual expectations

The previous sections have highlighted that structured environments provide an important source of spatial and content-based expectations that can shape attentional guidance. Only little is known about whether and how content-based preparatory activity is shaped by these expectations. A recent study (Gayet & Peelen, 2022) investigated whether preparatory activity adjusted to changes in an object's retinal size based on distance (Figure 4). Placing the same object at different distances from the observer creates vastly different retinal projections, but the relationship between retinal size is systematic and therefore predictable. Behavioural evidence indicates that this can be accounted for in search, showing that attentional capture is dependent on whether a presented object matches the predicted retinal size implied by distance cues in the scene (Gayet et al., 2024), and participants missing even large objects when their size is inconsistent with the surrounding scene (Eckstein et al., 2017). Decoding preparatory activity in visual cortex while participants prepared to search either near or far away in the scene showed that the representation of the search target in object-selective cortex was rescaled based on distance cues implied by the scene (Gayet & Peelen, 2022). This indicates that the preparatory attentional template can incorporate contextual expectations based on long-term experience.

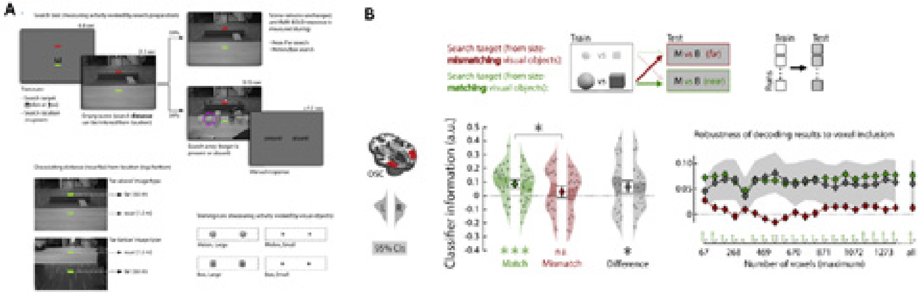

Figure 4: Preparatory activity adapts to scene context. A Search task of Gayet & Peelen (2022). Participants were searching for either melons or boxes at two distances in the scene: either near or far away, thereby changing the expected retinal size of the target object. Preparatory activity was measured on catch-trials in which participants fixated at a particular distance in the scene and prepared to search, but no search scene appeared. **B** Within object-selective cortex (OSC), decoding the current search target (melon or box) from preparatory activity with classifiers trained on visually presented large or small objects was better when classifiers were trained on isolated objects of matching than mismatching size, indicating that preparatory activity reflected a rescaled object-representation. This difference between matching and mismatching objects was not found in early visual cortex.

Aim and outline of this thesis

The studies presented in this thesis investigate the dynamic interplay between internal representations in visual cortex and the external world. The first two empirical chapters are focused on investigating how regularities in the external world are learnt and used for context-based attentional guidance and search, and how they shape attentional templates as well as preparatory activity. Finally, the last empirical chapter investigates whether such internal top-down biases during imagery directly compete with representations of external objects.

In Chapter 2, we test the flexibility and context-dependency of attentional templates for complex shapes retrieved from memory. Across a series of behavioural experiments, we find that different feature-dimensions of the same target shape can be flexibly emphasised, based on distractor expectations. Importantly, this adaptation of the template can be based on short-term selection history (priming) without explicit strategy or awareness, but also based on trial-by-trial context cues when the distractor contexts were separated in space, providing evidence for context-dependent attentional templates.

In Chapter 3, we test the functional role of preparatory activity as an internally-generated guiding template using fMRI and concurrent eye tracking. In line with the emerging view that attentional guidance is not necessarily based on a veridical representation of target features, we find that preparatory activity in the lateral occipital cortex (LOC) and intraparietal sulcus (IPS) reflects attention-guiding anchor objects, rather than the current target itself. Using context-dependent target-anchor associations, we dissociate features for guidance from target- and cue-related processing, and show how preparatory activity can support context-guided search based on object associations.

In Chapter 4, we test how externally-driven bottom-up signals interact with internally generated representations in visual cortex. Using EEG, we investigate the object representations of intervening objects while participants are engaged in a mental imagery task. By varying the similarity of imagined and perceived content, we test for competitive interactions at the category- and domain-level and find that concurrent imagery degrades representations of objects taken from the same category. This provides neural evidence for the hypothesis that internally and externally generated signals compete for representation within visual cortex.

Overall, this thesis sheds new light on the role and flexibility of internal representation in visual cortex, how they are shaped by the structure of the external world, and in turn shape what we perceive.

Chapter 2

Expected Distractor Context Biases the Attentional Template for Target Shapes

Visual search is supported by an internal representation of the target, the attentional template. However, which features are diagnostic of target presence critically depends on the distractors. Accordingly, previous research showed that consistent distractor context shapes the attentional template for simple targets, with the template emphasizing diagnostic dimensions (e.g., colour or orientation) in blocks of trials. Here, we investigated how distractor expectations bias attentional templates for complex shapes, and tested whether such biases reflect inter-trial priming or can be instantiated flexibly. Participants searched for novel shapes (cued by name) in two probabilistic distractor contexts: either the target's orientation or rectilinearity was unique (80% validity). Across four experiments, performance was better when the distractor context was expected, indicating that target features in the expected diagnostic dimension were emphasised. Attentional templates were biased by distractor expectations when distractor context was blocked, also for participants reporting no awareness of the manipulation. Interestingly, attentional templates were also biased when distractor context was cued on a trial-by-trial basis, but only when the two contexts were consistently presented at distinct spatial locations. These results show that attentional templates can flexibly and adaptively incorporate expectations about target-distractor relations when looking for the same object in different contexts.

This chapter is adapted from:

Lerebourg, M., de Lange, F. P., & Peelen, M. V. (2023). Expected Distractor Context Biases the Attentional Template for Target Shapes. *Journal of Experimental Psychology. Human Perception and Performance*.

Introduction

Goal-directed behaviour often requires visually searching for relevant objects in our environment. To find such target objects, we need to know the visual features characterizing them. For example, when searching for one's favourite green coffee mug among many other objects, we keep in mind the mug's specific colour and shape. The internal representation of the target is often called the *attentional template* or *search template* (Battistoni et al., 2017; J. Duncan & Humphreys, 1989; Eimer, 2014; Wolfe, 2021): a working memory representation of visual features that characterise the target. Attentional templates may act as a top-down bias (Desimone & Duncan, 1995), supporting search by guiding spatial attention to the target and/or supporting target identification (Hout & Goldinger, 2014; Wolfe, 2021).

To allow for efficient search, attentional templates should ideally include those features that distinguish the target from potential distractor objects (J. Duncan & Humphreys, 1989; Geng & Witkowski, 2019; Navalpakkam & Itti, 2007). Which features are diagnostic depends not only on the target itself but also on the current distractors: while looking for green objects may help to quickly find the green coffee mug on one's office desk, such a template would become much less useful if the mug was placed on the windowsill next to multiple green plants. Thus, when the distractor context is known (or predictable), this calls for a template that is shaped by expected distractors, flexibly highlighting diagnostic target features depending on context.

This idea has now gained some support. Generally, the statistical distribution of distractor features can be accurately learnt during search (Chetverikov et al., 2016, 2017). Based on this distractor knowledge, the template may be modulated accordingly, as evidenced by a range of studies investigating how attentional capture or memory representations of the target are influenced by predictable distractors. Within one specific feature dimension, such as colour or orientation, the attentional template could be shifted away from distractors (e.g. not the mug's particular shade of green but a slightly darker green, darker than all plants on the windowsill Geng, Di Quattro, et al., 2017; Kerzel, 2020; Scolari et al., 2012; Scolari & Serences, 2009; X. Yu & Geng, 2019a) or defined entirely relative to the distractors, without taking any specific feature value (e.g., a template for 'greener', but not a specific shade of green) (S. I. Becker, 2010b; S. I. Becker et al., 2010, 2013). These studies therefore provide evidence that attentional templates can be influenced by distractor context, allowing observers to distinguish targets from distractors more efficiently when these are characterised only by small feature differences. Context may affect processing at different search stages, and while there is evidence for context effects on attentional selection (S. I. Becker et al., 2010; Kerzel, 2020; Scolari & Serences, 2009) and sensory processing (Scolari & Serences, 2009), some of these effects could also reflect changes in later target identification and decision processes (Hamblin-Frohman & Becker, 2021; X. Yu, Hanks, et al., 2022).

Most targets and distractors are objects defined in multiple feature dimensions. Instead of shifting the template away from distractors within one dimension, features in more diagnostic dimensions (in which the target is unique, or nearly unique) may then also be emphasised in the template. In the coffee mug example, this could mean that the template more strongly reflects the mug's shape than the mug's colour when searching on the windowsill. Two recent behavioural studies provided evidence for this idea (Boettcher et al., 2020; J. Lee & Geng, 2020). In these studies, participants searched for gratings defined by orientation and colour (e.g., a blue grating at 45-degree orientation). In each block of trials, one dimension (e.g., colour, but varying particular features as red or blue across trials) could be used to differentiate the target from the distractors on most trials. On a subset of trials, however, the target was not unique in the expected dimension of that particular block, but instead stood out in the other dimension (e.g., orientation). Performance was worse on these mismatching trials, suggesting that target-features within the diagnostic dimension were emphasised relative to the expected undiagnostic one.

While these two studies provide evidence that attentional templates are influenced by distractor context, they leave several important questions unaddressed. First, while coloured gratings consist of two clearly separate dimensions, distinctly represented in early visual cortex, in daily life we typically search for object shapes that have a more unified representation in intermediate and higher-level areas of the ventral visual cortex (Kourtzi & Connor, 2011). Because most target objects are not defined by a combination of two distinct low-level features, it is not clear whether attentional templates for object shapes are similarly biased by distractor context to emphasise particular shape dimensions.

Second, it is unclear *exactly how context-dependent* such adaptive biases in the template are and what mechanism underlie these changes. Virtually all previous studies required participants to search repeatedly in contexts where either the distractor features (e.g., S. I. Becker, 2010b; Geng, Di Quattro, et al., 2017; X. Yu & Geng, 2019) or target-distractor relations (Boettcher et al., 2020; J. Lee & Geng, 2020) remained stable over an experimental block, or even the whole experiment. Expectations could therefore be based on recent selection history. Previous searches dynamically shape the attentional template, without necessarily requiring that target-distractor relations are learnt in a contextualised manner. Biases based on priming, a facilitation in performance based on repeating targets or distractors over successive trials, can reflect the statistics of targets and distractors (see Kristjánsson, 2022 for a review). Notably, such priming effects have been found for repetitions of specific features (e.g. Maljkovic & Nakayama, 1994), feature relations between targets and non-targets (e.g. spikier S. I. Becker, 2013), but also entire feature dimensions (e.g. overall colour; Found & Müller, 1996; Liesefeld et al., 2019). These biases are however typically short-lived. This is relevant as we often don't search repeatedly for the same object in the same context, but rather search for different objects in the same context or for the same object in different contexts. In such arguably more naturalistic situations (as in our coffee mug example), target-distractor relations would have to be learnt and bias the template in a context-dependent manner to account for changing target-distractor relations across contexts. Accordingly, it should be possible to base distractor expectations on long-term memory for different contexts, rather than short-term priming mechanisms. After learning these relations, re-encountering familiar contexts may flexibly shape the attentional template even on a trial-by-trial basis.

Finally, another relevant open question is whether the findings of previous studies relied on participants having explicit knowledge of the block structure or whether distractor context effects can arise implicitly. While Lee & Geng (2020) found equivalent effects of distractor context when participants were either explicitly informed about the distractor context manipulation or not, they did not assess whether participants were nonetheless aware of this in both cases. Considering that the stimulus dimensions in this experiment were clearly distinct and that the same dimension was diagnostic throughout half of the experiment, it is likely that participants were aware of the probabilities for different diagnostic dimensions and strategically used this. If templates are adapted to distractor context in daily life, this should occur relatively effortlessly and possibly implicitly.

Here, in a series of experiments, we addressed these open questions, probing when and how attentional templates are shaped by distractor context. We familiarised participants with novel, complex shapes and asked participants to search for them in different probabilistic contexts where one of the shape dimensions was diagnostic of the presence of the target. We first tested whether attentional templates for these shapes were biased by distractor expectations in a blocked context (Experiments 1 and 2) and assessing awareness of the manipulation (Experiment 2). Then, we investigated whether distractor expectations would also influence templates when context varied on a trial-by-trial basis (Experiments 3 and 4). To summarise the results, across experiments, we found flexible attentional templates for complex shape stimuli, shaped by distractor expectations in blocked contexts, independently of awareness. Further, we found evidence for context-dependent attentional templates, dynamically adjusted to context expectations on a trial-by-trial basis (Experiment 4).

Transparency and openness

Experimental stimuli, code and data are available on the project's OSF page (https://osf.io/tgu8r). All analyses were done using Matlab (2019a), the measures of effect size toolbox (Henschke, 2002) and JASP (JASP Team, 2022).

Design and analyses of Experiments 2B, 3 and 4 were preregistered on AsPredicted: https://aspredicted.org/bg5qa.pdf(Exp2B),https://aspredicted.org/e9zs9.pdf(Exp3), https://aspredicted.org/y4tq4.pdf (Exp 4).

To determine sample size, a priori power analyses were run using G*Power (Faul et al., 2007).

Experiment 1 – Diagnostic target dimensions are emphasised in blocked distractor contexts

In the first experiment, we tested the effect of blocked distractor context in which one target dimension was more discriminative than another when searching for complex 2D shapes. We created shapes varying along two dimensions: rectilinearity and orientation, jointly contributing to the outer form of the shape (see Figure 1)[1]. In addition, to increase similarity to real-world objects, all shapes were given names and additional elements that made them unique, characterizing them as individuals. While these stimuli were still controlled and simpler than real-world objects, they were designed to evoke an integrated representation. After familiarizing themselves with the target shapes and learning their names, participants searched for these shapes in displays in which either the target's rectilinearity or orientation was unique. To create predictable distractor contexts, unique rectilinearity and orientation trials were grouped in rectilinearity or orientation context blocks (with 80% of trials within a block belonging to this trial type). Participants were not explicitly informed about this. Due to the blocked structure, we could compare performance for the same trial type when it either matched or mismatched distractor expectations, based on the block they were presented in. Importantly,

---

[1] While rotation generally preserves shape, here orientation may also be seen as elongation, and our vertically or horizontally oriented shapes classified as either wide or long shapes.

displays on context-matching and -mismatching trials were visually identical, the only difference being the block context in which they appeared.

We chose to cue the shapes by name instead of presenting them visually prior to search. This required participants to rely on an internally generated memory representation based on the learnt associations between shape names and visual features to create a template. Arguably, generating a template in this top-down manner more closely resembles real-life search than visually cued search. However, compared to visually cued targets (e.g., J. Lee & Geng, 2020), this may lead to less flexible templates (see Boettcher et al., 2020 for a discussion). This is because, first, there is no external visual information of which different features or feature dimensions could be selectively encoded or maintained in visual working memory (e.g., Niklaus et al., 2017; Park et al., 2017), and, second, the memory representation on which the template is based has to be repeatedly retrieved across different contexts and may therefore not be biased by a particular context itself. Nevertheless, there is evidence that such associative templates can still be biased by context (Boettcher et al., 2020).

Methods

Participants

A total of 43 participants were recruited via Prolific (https://www.prolific.co) and took part in the online experiment. Thirty-four participants (mean age: 28.08, sd: 4.91; 22 female) were included in the analysis. Seven participants were excluded due to low task performance (< 55% accuracy, with chance level being 50%) and two further participants were excluded because they relied on a strategy based on shape familiarity without using the shape name cue in the search task: on trials where the cued shape was presented together with another distractor taken from the set of 12 target shapes on the other side of the display, their selection of the cued shape was not above chance (as determined by a binomial test with $\alpha = .05$). We excluded these participants because there was no evidence that the participants created a feature-specific template based on the cue (and which may then be modulated by distractor context expectations). Instead, they appeared to respond indiscriminately to any of the 12 familiar target shapes.

The final sample size of 34 was determined a priori. Assuming an effect size of about $\eta_p^2 = .20$ for the critical main effect of context match (a lower estimate consistent with results from J.Lee & Geng (2020)), this allowed for 80% power (with $\alpha = .05$) .

All participants provided informed consent and were paid £4.50. The study was approved by the Radboud University Faculty of Social Sciences Ethics Committee (ECSW2017-2306-517). All data were collected in 2021.

Procedure

The experiment was divided into a short shape-name training phase of at least three blocks (eight trials each) and six blocks (40 trials each) of the search task, taking around 35 minutes in total.

During the training phase participants learnt the names of 12 target shapes (see Figure 1), such that these names could later be used as cues in the search task. These individual shapes could be subdivided into four subgroups, sharing the same orientation (vertical or horizontal) and rectilinearity (rectilinear or rounded form). To support learning, all shape names provided information about the appearance of the shape, with the first letter (k or b) indicating its rectilinearity, the second letter (i or a) its orientation and the last two letters (ma, nu or so) the individual shape within that group. For each training block, participants first saw a display showing the four shapes and shape names belonging to the same individual (e.g., the baso, kaso, biso and kiso shapes). Then they completed eight self-paced trials in which one of the names was presented and the correct shape had to be selected out of the four individuals in the display. If six out of eight trials were answered correctly the next four individuals were presented, otherwise the training block was repeated.

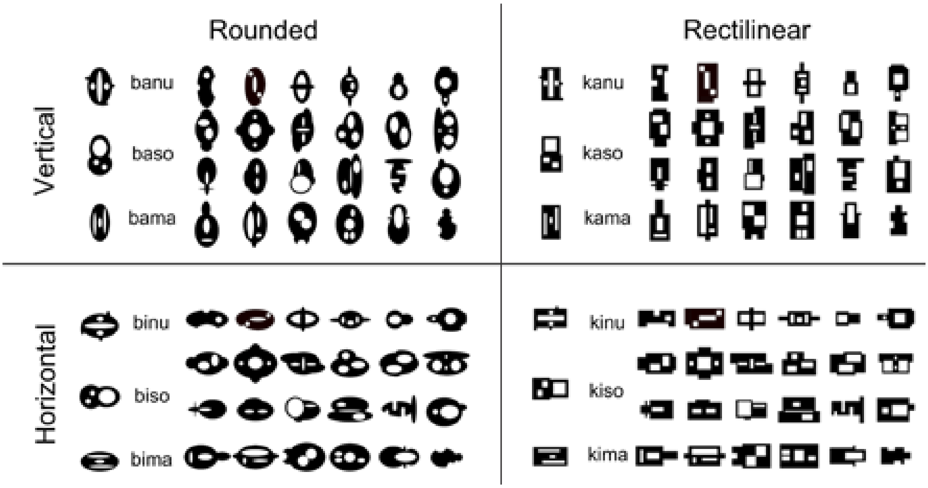

Figure 1: Overview of all target and distractor shapes. Target shapes are the named shapes in the leftmost column of each quadrant, all other shapes are distractors.

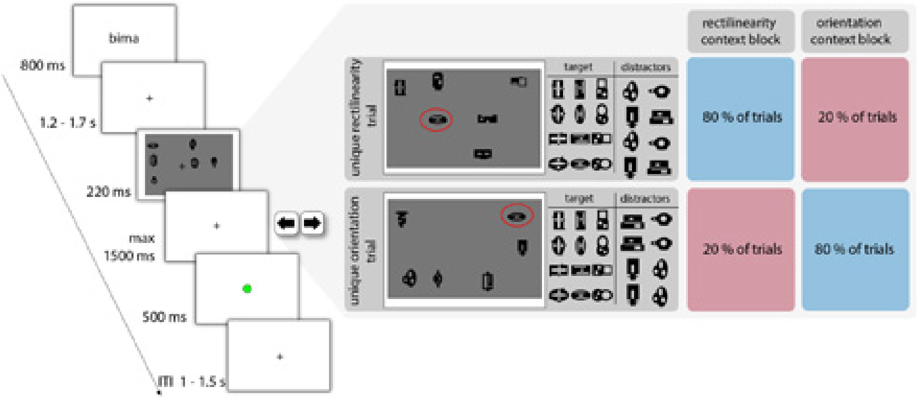

Figure 2: Timeline of a trial and design overview for experiment 1. At the start of each trial, participants saw a shape name cue, then a search display containing the target and other shapes appeared briefly and they indicated on which side the target had appeared. On half of the trials the target had a unique rectilinearity and on the other half a unique orientation. This discriminative dimension was blocked, with 80% validity. The same search displays were thus presented in blocks in which their diagnostic dimensions matched (red) or mismatched (blue) the current block context. The target is highlighted in the example displays for illustration in this figure only. Overall, all distractors were equally frequent in both types of trials, but paired with different targets (only some example distractors are shown here).

After successfully completing training, participants proceeded to the search task (see Figure 2). At the start of each trial, one of the 12 target shapes was cued by name. After a delay period, a search display consisting of six shapes appeared for 220 ms. Participants had to report on which side of the display the target had appeared by using the left and right arrow keys of their keyboard, before receiving feedback via a coloured dot (either red or green) appearing at fixation. From the perspective of the participant, the task was always to find the cued shape. What was not mentioned to them was that on every trial, the target was either unique in its rectilinearity (e.g., a vertical rectilinear shape among rounded vertical and horizontal shapes) or orientation (e.g., a vertical rectilinear shape among horizontal rectilinear and rounded shapes) from all other distractors. In total, there were 120 unique orientation and 120 unique rectilinearity trials. Trial type was blocked with 80% validity, yielding rectilinearity and orientation context blocks (with 32 trials within a block of one trial type and eight of the other). Mismatching context trials in the rectilinearity context block therefore had a unique orientation and vice versa. Importantly, within a block, there was not a specific feature (e.g., a horizontal shape) that was more common for either target or distractors, as all vertical and horizontal, rectilinear and rounded shapes were overall equally frequent. The difference between blocks was which targets and distractors were paired together in specific displays, making a dimension (e.g., orientation) but not any specific feature (e.g., horizontal) diagnostic. This design avoided low-level adaptation and intertrial priming of specific features (Maljkovic & Nakayama, 1994).

Target orientation and rectilinearity were counterbalanced within each block, while individual target shapes and the side on which the target appeared were counterbalanced across all 240 trials of the experiment. Block type (rectilinearity or orientation context blocks) alternated and the first 8 trials of a block were always valid (i.e., belonging to the dominant trial type of that block). The initial block context (rectilinearity or orientation) was randomised.

Stimuli & Setup

A total of 108 novel, black shapes were created using Adobe Illustrator (see Figure 1). These shapes could be grouped into four categories based on their rectilinearity (rectilinear or rounded base shape) and orientation (upright or horizontal), comprising 27 individuals each. All individual shapes had a different configuration of local elements (e.g., small holes/protrusions of varying sizes and positions) and also differed slightly in their size and width-height ratio. The arrangement of these configurations was kept across shape groups, meaning that highly similar individual shapes appeared in all 4 shape groups. Shapes were created by adding these local elements to an oval or rectangular base shape, exchanging all rectangles by circles or ovals and/or rotating the resulting shapes to create similar shape exemplars across groups. Three shapes of each group were chosen as target shapes.

To create search displays, five distractor shapes and the target shape were placed on a grey background aligned on an invisible 3 x 4 grid, with the constraint that three shapes should appear on each side of the fixation cross (Figure 2). Random jitter (+/- 0.2 x the shape size) was added to the grid positions. In each display, distractors were chosen such that the target was always unique in one of its dimensions but shared features with some distractors in the other dimension, to reduce pop-out of the target. Of the five distractors, three had the same feature value (e.g., horizontal orientation) of the target in the non-discriminative dimension. Two distractors did not share any feature with the target and were presented on the same display side as the target. This allowed to distinguish responses to the target and to the most similar distractors. For a given target shape, distractors in the display could also include other target shapes from other shape groups, but these were not more likely than any other non-target distractor.

The experiment was programmed in Psychopy (Peirce et al., 2019) and hosted on Pavlovia (https://pavlovia.org). All participants took part in the experiment using their own computer. At the beginning of the experiment, their screen resolution was determined by asking them to rescale a credit card picture on the screen to the physical size of their credit card and they were asked to sit at about an arm's length (~57 cm) from their screen (Q. Li et al., 2020). Assuming this viewing distance, the search displays subtended around 20 x 15°.

Data Analysis

For our analyses, accuracy and reaction were combined into a single measure, the linear integrated speed accuracy score (LISAS; Vandierendonck, 2017). This provided us with a single performance measure that is sensitive to effects in accuracy and RT in the same direction, accounting for additional variance in performance compared to its two individual components. Individual results for accuracy and RT for every experiment are included in the Appendix. Additional figures for accuracy and RT results can also be found in the supplementary materials (figures S1 – S4).

The LISAS is a linear combination of RT and error rates, the latter scaled by the standard deviation of both measures, thus increasing reaction times as a function of error rates. It is defined as: $LISAS = RT_{ij} + PE_{ij} \times \frac{S_{RT_i}}{S_{PE_i}}$, with RT_{ij} being the mean (correct) response time of participant i in condition j, PE_{ij} participant i's error rate in condition j, S_{RTi} the standard deviation of response time (for participant i) and S_{PEi} the standard deviation of responses (coded as 0's and 1's for correct/incorrect) of this participant.

To test for effects of distractor context expectations, we analysed performance on the unique rectilinearity and orientation trials as a function of the block context using a 2 (trial type: unique rectilinearity, unique orientation trial) x 2 (context match: presented in matching, mismatching context block) repeated-measures ANOVA.

Context matching effects for individual dimensions were followed up with t-tests (2-sided, α = .05). Bayes factors for simple t-contrasts and the main effects/interactions were calculated as Bayesian one-sample t-tests for the respective difference scores between conditions or condition averages using the JASP default settings (Cauchy prior with scale 0.707).

Results

Trials with reaction times below 150 ms and those +/- 3 standard deviations away from participant's mean RT on correct trials were excluded, resulting in rejection of 1.26% (sd 0.93) of trials.

Overall accuracy was 82.25% (sd 8.52), overall RT was 696.83 ms (sd 82.85).

As hypothesised, participants performed better when distractor context was expected, i.e. when a unique rectilinearity/orientation trial appeared in the matching compared to mismatching block context (Figure 3; main effect of context match: $F(1,33) = 7.84$, $p = .009$, $\eta_p^2 = .19$, $BF_{10} = 4.98$).

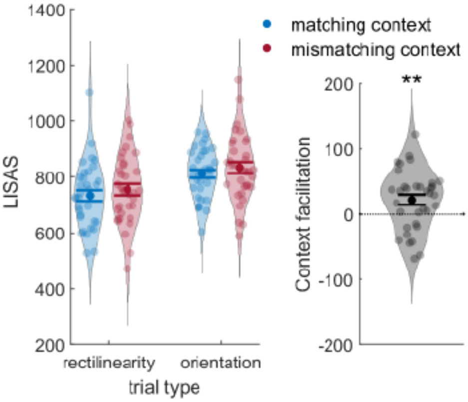

Figure 3: Results of Experiment 1. Performance on both trial types as a function of block context (left) and context facilitation (LISAS mismatching – matching context) averaged across trial types (right). All error bars are SEM.

In addition, participants were overall better on the unique rectilinearity trials (main effect of trial type: $F(1,33) = 50.92$ $p < .001$, $\eta_p^2 = .61$, $\mathrm{BF}_{10} > 3 \times 10^5$). This suggests that the target was generally easier to find when it had a unique rectilinearity than when it had a unique orientation. However, context expectations influenced search equally for both trial types (no significant interaction between trial type and context match: $F(1,33) = 0.00$, $p = .99$, $\eta_p^2 < .001$, $\mathrm{BF}_{10} = 0.18$; context match rectilinearity unique trials: $t(33) = 1.51$, $p = .14$, $d_z = 0.26$, $\mathrm{BF}_{10} = 0.51$; orientation unique trials: $t(33) = 1.43$, $p = .16$, $d_z = 0.24$, $\mathrm{BF}_{10} = 0.46$).

Discussion

Experiment 1 showed that search performance was influenced by distractor context expectations. Searching for the same target shapes in different block contexts, participants performed better when the diagnostic target dimension matched their expectations compared to when it did not, even though the search displays were identical in these conditions. These findings provide an important extension of previous work (Boettcher et al., 2020; J. Lee & Geng, 2020) , showing that probabilistic distractor context influences search performance for complex shapes with individual features and names. Notably, these biases also arose relatively quickly, as each block only had 32 valid trials and learning of distractor context could only be based on briefly flashed displays with heterogeneous distractors differing in the non-diagnostic dimension.

We interpret these results as reflecting a change in the attentional template, with a different template being used for the same object in different contexts based on distractor expectations. As we used name cues, this bias cannot be explained in terms of selective encoding or maintenance (Niklaus et al., 2017; Park et al., 2017) of external visual information into working memory. Thus, we show that internally-generated templates for complex shapes, as used in naturalistic search, are flexibly shaped by distractor expectations.

Experiment 2 – Probing biases in the target template

Our interpretation of the results of Experiment 1 is that the attentional template for the shapes was modulated by distractor expectations, and that the target feature in the relevant dimension was thereby emphasised over the irrelevant dimension (J. Lee & Geng, 2020). However, another possibility might be that this representation of the target shape itself did not change, but rather that participants were learning about some property of the entire display, without changing their template for the target shape as such.

For example, despite attempts to make the displays homogeneous and the target not pop out, our target shape was by definition still a singleton in one dimension. Participants could therefore have learnt to look for orientation or rectilinearity singletons depending on the block and not created/modulated shape-specific target templates. This is especially relevant given that singleton search is facilitated when the target-defining dimension repeats across trials (e.g. Found & Müller, 1996). Indeed, according to the dimension weighting account of attention (see Liesefeld et al., 2019 for a review), feature dimensions can be weighted differently when information is integrated across dimensions into a single spatial priority map (however see S. I. Becker, 2010a; S. I. Becker et al., 2014 arguing for a post-selectional account for dimension-weighting effects in these studies). These effects are however different from a change in the attentional template to emphasise a specific feature, as singleton search only requires to find the odd-one out without requiring a specific template. Because this account is based on saliency or contrast, dimension weighting also applies equally to all features in one dimension (see also J. Lee & Geng, 2020).

Along similar lines, based on which distractors appeared together in displays of different contexts, participants could also have learnt to efficiently group and suppress distractors with a common rectilinearity in the rectilinearity context blocks and a common orientation in the orientation block.

To test whether our predictive distractor context indeed changed the shape-specific attentional template, we adapted our paradigm to include 2-shape probe-trials, in which the target was presented along with the same individual of another shape group (shapes sharing similar local elements, but always differing in one dimension, e.g., a kaso and a kiso shape). These trials were presented intermixed with the regular 6-shape search displays, for which the discriminative dimension was always consistent within a block. If the target template was indeed shaped by context, selecting the target should be easier based on the discriminative dimension of the 6-shape displays in this block. We would therefore expect to replicate the results of Experiment 1 with the two-shape displays. If, instead, participants learnt to select a specific singleton or to suppress certain distractor groups in the display, this learning should not transfer to the 2-shape trials and performance on these trials should therefore be independent of context.

In addition, we were also interested in whether our results reflected explicit knowledge of distractor context. Previous studies provided evidence that different aspects of distractor context, including co-occurrence between distractors (Thorat et al., 2022), their spatial arrangements in contextual cueing paradigms (Chun & Jiang, 1999; Spaak & de Lange, 2020), and the shape of their underlying feature distribution (Hansmann-Roth et al., 2021) can be learnt implicitly, without awareness of these regularities. Here, we introduced two questions with ratings at the end of the experiment, asking participants whether they had become aware of the distractor context manipulation and whether they had used it strategically. We then related these ratings to context facilitation.

Methods

Participants

Data from 34 participants (18 female; mean age: 28.44, sd: 4.93) were included for the original experiment. Two participants were replaced due to relying on a familiar shape strategy rather than using the shape name cue, as described in Experiment 1. As for Experiment 1, sample size was determined a priori and allowed to detect a main effect of context match with effect size of $\eta_p^2 = .20$ with 80% power ($\alpha = .05$).

After the original experiment (Experiment 2A), we ran a higher-powered, exact replication (Experiment 2B). Eighty new participants were included (45 females, one person who did not indicate their gender; mean age: 29.18, sd: 4.12). Five participants were replaced due to overall low accuracy (<55%) and an additional one because they did not use the shape name cue. Based on this sample size, we were able to detect a smaller effect (about half of the effect size in Experiment 1) with 80% power.

All participants were recruited via Prolific (https://www.prolific.co), provided informed consent prior to participation and received £5.00. The study was approved by the Radboud University Faculty of Social Sciences Ethics Committee (ECSW2017-2306-517). All data were collected in 2022.

Procedure

Participants were first familiarised with the shapes and their names using the same training procedure as in Experiment 1. Afterwards, they completed ten self-paced practice trials and six blocks (64 trials each) of the actual search task.

On each trial, participants saw a shape name cue, followed by a search display containing either six or two shapes. Regardless of the number of shapes shown, participants always indicated on which side of the display (left or right) the target shape had appeared. Six-shape displays appeared on three out of five trials (40 trials per block) and within a block, all had the same discriminative dimension, creating the block context in which the two-shape displays were presented (Figure 4). The two-shape displays contained the target shape as well as the same individual shape from another target group (i.e., differing from the target shape in either rectilinearity or orientation, but sharing the other dimension and having similar other local elements) on the other side of the fixation cross. Independent of the block context, on half of these two-shape trials, the target shape had a unique rectilinearity; on the other half, it had a unique orientation. Collapsed over six- and two-shape trials, the diagnostic dimension was 81.25% valid within a block, similar to Experiment 1.

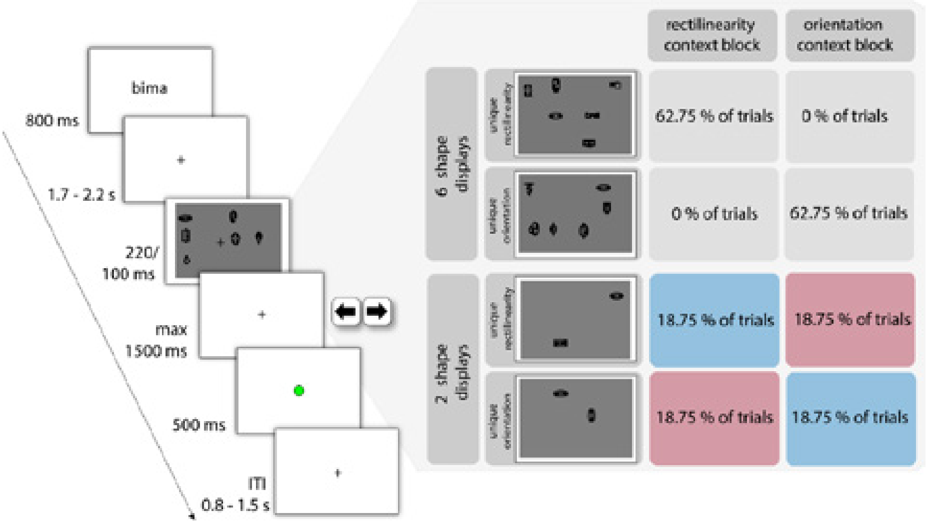

Figure 4: Timeline of a trial and design of Experiment 2A and 2B. Shape targets were cued by name and had to be searched for in displays of 2 or 6 shapes. Six-shape displays within one block all had the same diagnostic dimension and created the context in which two-shape displays were presented. Those always contained the target and the same individual taken from another shape group, differing either in rectilinearity or orientation and therefore matching (blue) or mismatching (red) the current context.

In an attempt to equalise difficulty across display types, two-shape displays were presented for 100 ms only, compared to 220 ms for the six-shape displays. Following their response, participants received feedback via a red/green circle presented at fixation.

After completing the experiment, participants were told about the distractor context manipulation and asked to report 1) whether they had noticed that all 6-shape trials in a block had a unique discriminative dimension, and 2) whether they had actively used this, by providing ratings on five-point scales.

Target orientation and rectilinearity was counterbalanced within blocks, while individual target shapes and display side of the target were counterbalanced across the whole experiment. All factors were counterbalanced separately for each display type. Block context alternated and the block context participants started with was randomised.

Stimuli & Setup

The 6-shape displays were created as in Experiment 1. For the two-shape displays, the target shape was placed on one side of the display, and the same-individual shape from another shape group (sharing either rectilinearity or orientation and other local elements with the target) on the other side. As for the six-shape displays, an invisible 3 x 4 grid was used to place the shapes.

All participants took part in the experiment on their own device. The experiment was coded using Psychopy and hosted on Pavlovia (https://pavlovia.org).

Data Analysis

As for Experiment 1, we conducted a 2 (trial type: unique rectilinearity, unique orientation trial) x 2 (context match: presented in matching, mismatching context block) repeated measures ANOVA on the LISAS of the two-shape trials.

In addition, we compared the effects of context match on the two-shape trials of Experiment 2B with the context match effects of Experiment 1 using a 2 (trial type: unique rectilinearity, unique orientation trial) x 2 (context match: presented in matching, mismatching context block) x 2 (Experiment: Experiment 1, 2B) repeated measures ANOVA, with Experiment as between participant factor.

Context matching effects for individual dimensions were followed up with t-tests (two-sided, $\alpha = .05$). Bayes factors for simple t-contrasts and the main effects/interactions were calculated as Bayesian one-sample t-tests for the respective difference scores between conditions or condition averages using the JASP default settings (Cauchy prior with scale 0.707). For comparisons between experiments or groups, Bayesian independent sample t-tests were computed instead.

Results

Experiment 2A

We excluded trials in which response times were below 150 ms or +/- 3 standard deviations away from the participant's mean correct RT for the respective display type, on average leading to a rejection of 1.56% (sd 0.87) of trials across display types.

Accuracy for the six-shape trials was 89.35% (sd 6.97) and mean correct RT was 726.37 ms (sd 84.44). For the two-shape trials, accuracy was 90.52% (sd 8.61) and mean correct RT was 670.18 ms (sd 84.48).

All subsequent analyses focused on the two-shape trials, as the six-shape trials did not include context mismatch trials.

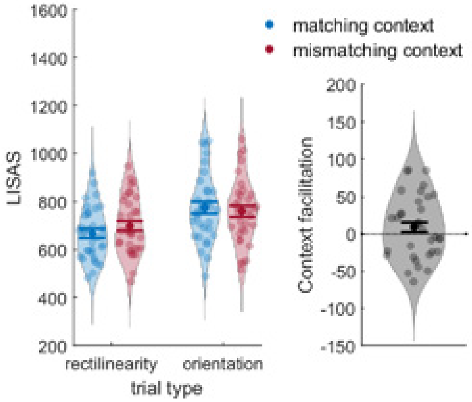

Figure 5: Results of Experiment 2A. Performance on the two-shape trials for both trial types as a function of block context (left panel) and context facilitation (LISAS mismatching—matching context) averaged across trial types (right). All error bars are SEM.

Overall, there was a slight improvement in performance when the target's diagnostic dimension was expected based on block context, but this was not significant (Figure 5; main effect of context match: $F(1,33) = 1.78, p = .19, \eta_p^2 = .05, BF_{10} = 0.42$). As previously, trials in which the target's rectilinearity was diagnostic were generally easier (trial type: $F(1,33) = 34.77, p < .001, \eta_p^2 = .51, BF_{10} > 1000$). In this experiment, the effect of context expectations also differed between trial types (context match x trial type: $F(1,33) = 7.39, p = .01, \eta_p^2 = .18, BF_{10} = 4.20$): for rectilinearity unique trials performance improved when they were presented in a matching context ($t(33) = 2.64$, $p = .01$, $d_z = 0.45$, $\mathrm{BF}_{10} = 3.54$), but this was not the case for unique orientation trials ($t(33) = -1.43$, $p = .16$, $d_z = -0.25$, $\mathrm{BF}_{10} = 0.46$).We then tested for overall awareness of our context manipulation and its relation to context facilitation. Twenty-eight participants provided an awareness rating, ranging from 1 (I didn't notice this at all) to 5 (it was very obvious). The mean value was 3.00 (sd 1.39), with 6 participants indicating they had not noticed the relevant dimension being consistent within blocks at all (corresponding to a rating of 1) and 12 participants giving it a rating >= 3.

For the second rating, regarding whether participants actively made use of context knowledge, we excluded participants who had responded with 1 for the awareness rating (as they could not provide meaningful answers to this question), resulting in 22 ratings. The rating ranged from 1 (I didn't use this at all) to 5 (I used this a lot and prepared differently depending on the blocks). The mean rating was 2.95 (sd 1.39), with three participants indicating they had not made use of it at all (rating of 1) and 12 providing a rating >= 3.

We correlated both of these ratings with participant's context facilitation (LISAS mismatching – matching context, averaged across both trial types). Context facilitation did not correlate with either the awareness (ρ = - .08, p = .69) or the usage rating (ρ = -.13, p = .56). This was also true when correlating only context facilitation in unique rectilinearity trials with awareness (ρ = .02, p = .91) or usage ratings (ρ = -.13, p = .56).

Experiment 2B

We excluded anticipatory and delayed responses for every participant, as preregistered and described for the original experiment, resulting in an average loss of 1.54% (sd 0.77) of trials.

For the six-shape trials, accuracy was 89.16% (sd 7.47) and RT was 731.06 ms (sd 101.02). For the two-shape trials, accuracy was 90.79% (sd 8.28) and RT was 666.48 ms (sd 96.67).

The analysis was the same as Experiment 2A. In line with our hypothesis, we found that performance improved when the diagnostic dimension matched participants' expectations (Figure 6; main effect of context match (F(1, 79) = 9.98, p = .002, η_p^2 = .112, BF_{10} = 11.70). Our higher-powered replication thus confirmed that the target template was indeed modulated by distractor expectations.

Again, performance was overall higher when rectilinearity was the diagnostic target dimension (trial type: F(1,33) = 95.76, p < .001, η_p^2 = .55, BF_{10} > 2x 10 ^{12}). Unlike Experiment 2A, the effect of context expectations was independent of which dimension was diagnostic (context match x trial type: F(1,33)= 0.28, p = .60, η_p^2 = .00, BF_{10} = 0.14; context match unique rectilinearity trials: t(79) = 1.32, p = .19, d_z = 0.15, BF_{10} = 0.28, unique orientation trials: t(79) = 2.13, p = .04, d_z = 0.24, BF_{10} = 1.04).

We also tested whether context facilitation on the two-shape trials differed from the one found for 6-shape trials in Experiment 1 in an exploratory analysis. Even though these differed in some aspects (most importantly that the two-shape trials were generally easier, but also the number of trials), context facilitation did not differ across experiments (main effect of experiment: F(1,112) = 7.68, p = 0.007, η_p^2 = .06, BF_{10} = 6.08; no interaction between context match x Experiment: F(1, 112) = 1.10, p = .30, η_p^2 = .01, BF_{10} = 0.26).

In addition to the main preregistered analysis, awareness and usage ratings were also collected for this experiment. We collected 76 awareness ratings, having a mean value of 2.34 (sd 1.34), with 29 subjects indicating that they had not noticed this at all (giving a rating of 1) and an equal number of subjects giving it a rating >= 3. For the usage rating, 47 ratings were included and the mean value was 2.89 (sd 1.27), with 28 ratings >= 3.

As in Experiment 2A, neither the awareness, nor the usage rating correlated with the context facilitation across participants (awareness rating: $\rho = .08$, $p = .48$, usage rating: $\rho = -.09$, $p = .55$), suggesting that context facilitation was not related to explicit knowledge about the distractor context.

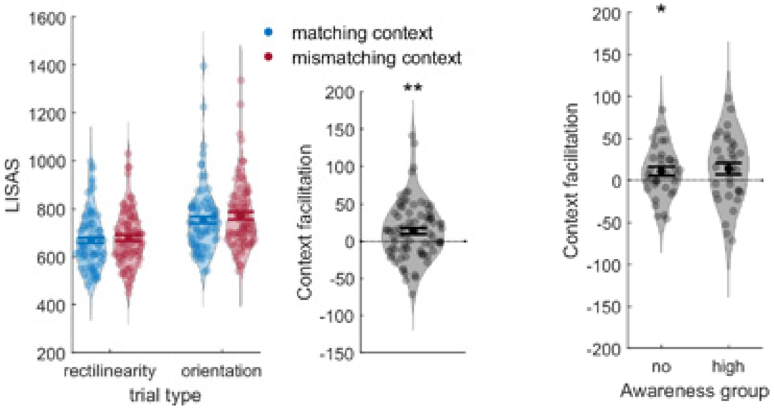

Figure 6: Results of Experiment 2B. Performance on the two-shape trials for both trial types as a function of block context (left panel) and context facilitation (LISAS mismatching—matching context) averaged across trial types (middle panel). The right panel shows context facilitation for Exp 2A and 2B combined, split by participants' reported awareness of the context manipulation. All error bars are SEM.

To test this more explicitly, we pooled data from Experiment 2A and 2B to repeat our analysis of distractor context effects separately for participants who reported not having noticed the distractor manipulation (awareness rating = 1) and those who found it relatively obvious (awareness rating >= 4) in an exploratory analysis. Most importantly, we found that in the no-awareness group, performance still improved when the diagnostic dimension was expected based on context (main effect of context match; no awareness group: $F(1,33) = 4.53$, $p = .04$, $\eta_p^2 = .12$, $BF_{10} = 1.33$; high awareness group: $F(1,34) = 4.00$, $p = .05$, $\eta_p^2 = .11$, $BF_{10} = 1.07$), and the effect of context expectations was not stronger for the high awareness group ($t(67) = -0.31$, $p = .75$, $d_z = 0.05$, $BF_{10} = 0.26$). Also the correlations of context facilitation with awareness and usage ratings were not significant across this larger sample (awareness rating: $\rho = .01$, $p = .96$, usage rating: $\rho = -.13$, $p = .29$).

Discussion

Experiment 2 provided additional evidence that distractor expectations shapes the representation of target features: selecting the target in the two-shape trials based on its rectilinearity was easier in the rectilinearity blocks than orientation context blocks and vice versa. By intermixing two-shape trials in the blocks we were able to show that distractor context affected the attentional template, and therefore the representation of the target shape, rather than more efficient grouping of distractors or singleton detection.

In Experiment 2A (but not Experiment 2B), we observed a difference in context facilitation across diagnostic dimensions, with the modulation by distractor expectations being stronger for the rectilinearity than the orientation dimension. A similar kind of asymmetry was found in previous studies (Boettcher et al., 2020; J. Lee & Geng, 2020), where colour was more consistently modulated by context compared to orientation. One interpretation is that feature dimensions which generally allow for easier discrimination (and the target can be more easily selected on their basis, as rectilinearity in our task, or colour in others) are more accessible and likely to be further emphasised or deemphasised based on context.. In addition, if distractors are sharing target features in this dimension, they may be particularly distracting and require adjusting the template, more so than for overall less prominent dimensions. For example, if target rectilinearity is more strongly represented than orientation in the template by default, especially the distractors sharing target rectilinearity in the unique orientation trials would hamper performance (in line with the overall worse performance observed on these trials). Encountering many such distractors in the unique orientation blocks would require adjusting this default, and decreasing the reliance on rectilinearity. Therefore performance on the rare unique rectilinearity trials in orientation blocks may be worse than in the rectilinearity unique blocks, leading to a context facilitation effect for rectilinearity trials.

Was this change in the template due to deliberate and explicit processes or could it occur more implicitly? The awareness ratings suggest that our manipulation became noticeable to several participants at some point during the experiment, and a subset of those also reported actively using knowledge of the likely discriminative dimension. However, both ratings did not correlate with the magnitude of context facilitation across participants. Further, we also found context facilitation for those participants who reported not having noticed any distractor regularities, which was equal in magnitude to the facilitation shown by participants who did notice them. While this small effect should be interpreted with care and further replicated, modulation of the attentional template by distractor expectations may therefore not depend on participants noticing and deliberately using distractor regularities, similar to the implicit learning of other types of distractor regularities (Chun & Jiang, 1999; Hansmann-Roth et al., 2021; Spaak & de Lange, 2020; Thorat et al., 2022).

Experiment 3 – Adjusting to distractor expectations on a trial-by-trial basis

Having established that expected discriminative target features could be emphasised in the attentional template based on a blocked context, in Experiment 3 we tested whether the attentional template is also updated when distractor contexts are familiar but changing frequently, as would be necessary in many daily-life situations (e.g., searching for your keys in different rooms or on different tables). This would be evidence for a flexible top-down bias based on long term memory for respective contexts, independent of priming mechanisms.

In this way we could test whether target-distractor relations were learnt separately for different contexts and thus whether distractor expectations modulated search performance in a truly context-dependent manner. Previous research has shown that some statistics of targets or distractors, e.g. recent colour statistics within object categories (Kershner & Hollingworth, 2022) or features associated with rewards (Anderson, 2015) are learnt and used in that way, with context cues modulating the attentional template on a trial-by-trial basis. Furthermore, there is also neural evidence that context cues can be used to tune sensory gain of orientation-selective neurons in visual cortex away from distractor orientations on a trial-by-trial basis (Scolari et al., 2012).

To provide context cues, we paired the two different contexts (either unique rectilinearity or orientations as frequent diagnostic dimension) with two distinct background colours. To facilitate learning of the colour-context associations, the distractor context was blocked for the first half of the experiment (as in Experiments 1 and 2). Afterwards, distractor context changed randomly on every trial.

Our design now required participants to not only learn regularities in diagnostic dimensions, but also to associate these with different colour cues, which is more challenging. Since the main focus of this study was not on whether these additional associations could be learnt implicitly but rather on whether they were being used after learning, we now informed participants about the meaning of those colour cues.

Methods

Participants

Eighty participants (52 female, one non-binary person; mean age: 28.63, sd: 4.17) were included in this experiment. Five additional participants were replaced due to low accuracy (<55%). We assumed the effect of context expectations may be reduced in the interleaved condition, and this sample size allowed us to detect main effect of context match effect about half as large as the blocked condition in Experiment 1 with 80% power ($\eta_p^2 > .095$, with $\alpha = .05$).

All participants were recruited via Prolific (https://www.prolific.com), provided informed consent and received £5.00 for their participation. The study was approved by the Radboud University Faculty of Social Sciences Ethics Committee (ECSW2017-2306-517). All data were collected in 2022.

Procedure

The overall procedure was similar to Experiment 1 (Figure 7). Participants first completed the same shape-name training before starting the search task. The search task was explained to them, including the distractor context manipulation and that the background colour of the search display provided information about the likely upcoming distractor context. Before the start of the search task, participants also completed ten self-paced practice trials.

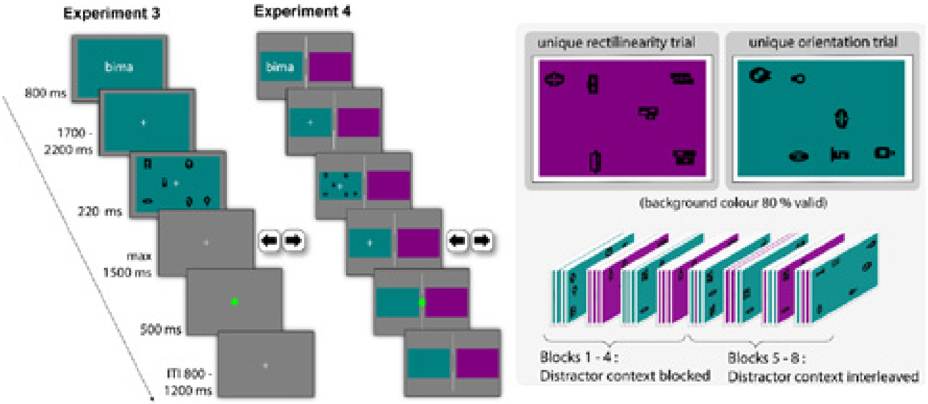

Figure 7: Timeline of a trial and design overview for Experiments 3 and 4. The overall search task was the same as in Experiment 1, but the background colour predicted the diagnostic target dimension with 80% validity. Distractor context was blocked for the first half of the experiment and interleaved for the remaining half. In Experiment 3, background colour changed at the onset of a trial. In Experiment 4, both contexts remained constantly on screen, on the left and right of the display.

The search task consisted of eight blocks of 40 trials each. To facilitate learning, in the first four blocks, background colour within a block was constant (alternating between blocks), indicating the upcoming trial type with 80% validity. In the second half of the experiment, background colour changed on a trial-by-trial basis, again indicating the upcoming trial type with 80% validity. This design should minimise interference between the two contexts during learning, allowing to establish separate representations and learning rules for both (Flesch et al., 2018). Background colour changed as soon as the name cue was presented on screen and remained visible until the search display had disappeared. Reminders of the associations between distractor contexts and background colour were also shown at the beginning of every block. The experiment took around 50 minutes in total.

Target orientation, rectilinearity and validity were counterbalanced per block, while the specific target individuals and the display side on which the target appeared were counterbalanced across the whole experiment. All factors were counterbalanced separately for both contexts. Which block context participants started with and the associations between block context and background colour were randomised.

Stimuli & Setup

Stimuli were created as in Experiment 1, with the exception that the background colour was changed to either blueish green (RGB 0, 127.5, 127.5) or magenta (RGB 127.5, 0, 127.5). The experiment ran online, was programmed in Psychopy and hosted on Pavlovia (https://pavlovia.org). Participants used their own devices to take part in it.

Data Analysis

We analysed distractor context effects separately for the blocked and interleaved context blocks with two 2 (trial type: rectilinearity, unique orientation trial) x 2 (context match: matching, mismatching context cue) repeated measures ANOVAs.

To understand how context facilitation evolved within blocks, we compared context facilitation across different trial bins in the blocked context pooled data from all 6-shape trials (all six blocks from Experiment 1 and the first four blocks of Experiment 3. We computed context facilitation by computing the performance difference between matching and mismatching context trials in each time bin, independent of dimension, and analysed it in a 1 x 4 ANOVA (block time bin: trials 9-16, trials 17-24, trials 25 – 32, trials 32 - 40).

In addition, we tested the effects of intertrial priming on performance in the interleaved condition. For this, we computed performance measures separately for trials in which the diagnostic dimension was the same as on trial n-1 and trials for which the diagnostic dimension had switched from the previous trial, only including those trials on which the previous response had been correct, but independent of the context cues.

Context matching effects for individual dimensions were followed up with t-tests (two-sided, $\alpha = .05$). Bayes factors for simple t-contrasts and the main effects/ interactions were calculated as Bayesian one-sample t-tests for the respective difference scores between conditions or condition averages using the JASP default settings (Cauchy prior with scale 0.707). For comparisons between experiments, Bayesian independent sample t-tests were computed instead.

Results

Following our preregistration, we excluded trials in which reaction times were below 150 ms or +/- 3 std away from this participant's mean correct RT. This resulted in a rejection of 1.57% (sd 0.73) of trials.

Overall accuracy was 87.60% (sd 8.18) and mean RT was 702.74 ms (sd 82.40).

Replicating our previous experiments, context expectations facilitated performance when context was blocked (Figure 8; context match: $F(1,79) = 14.74$, $p < .001$, $\eta_p^2 = 0.157$, $BF_{10} = 78.04$). Context facilitation depended on which dimension was diagnostic (interaction context match and trial type: $F(1,79) = 10.95$, $p = 0.001$, $\eta_p^2 = 0.12$, $BF_{10} = 16.95$), matching context improved performance on unique rectilinearity unique trials ($t(79) = 5.08$, $p < .001$, $d_z = 0.57$, $BF_{10} > 5000$) but not unique orientation trials ($t(79) = -0.74$, $p = .46$, $d_z = -0.08$, $BF_{10} = 0.16$).

In the interleaved condition however, where context changed on a trial-by-trial basis, context cues did not influence search performance (main effect context match: $F(1,79) = 1.68$, $p = 0.20$, $\eta_p^2 = 0.021$, $BF_{10} = 0.28$), and this was equally the case for both diagnostic dimension (no interaction between context match and trial type: $F(1,79) = 0.12$, $p = .72$, $\eta_p^2 = 0.00$, $BF_{10} = 0.13$).

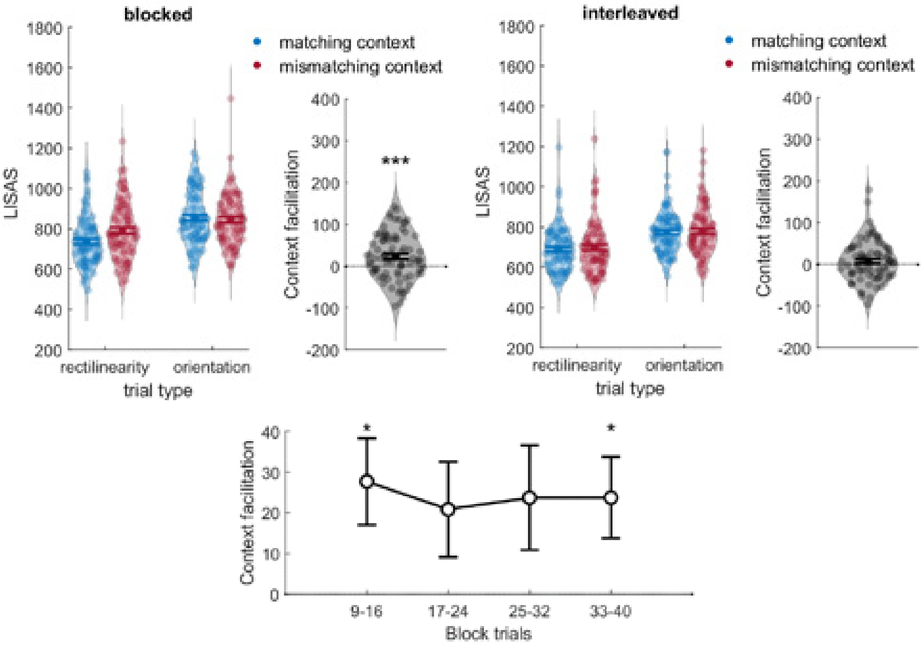

Figure 8: Results of Experiment 3. The upper panel shows performance on both trial types as a function of context (left subpanel) and context facilitation (LISAS mismatching—matching context) averaged across trial types (right subpanel). All error bars are SEM. The lower panel shows context facilitation calculated for individual trial bins of the blocked condition in Experiments 1 and 3 combined.

Our results suggested that context may need to be stable over trials in order to influence search performance, leading to two additional exploratory analyses. First, one may ask whether these context expectations could be established relatively quickly within the blocked conditions or gradually developed after multiple trials. We analysed context facilitation across 4 block time bins across pooled data from Experiment 1 and 3, excluding the first eight trials which were always valid. Generally, context facilitation remained stable across trial bins (main effect of trial bin: $F(3,336) = 0.01$, $p = .99$, $\eta_p^2 < 0.001$, $BF_{10} = 0.009$). In addition, context facilitation already arose in the earliest analysable time bin (trials 9-16, $t(113) = 2.57$, $p = .01$, $d_z = 0.24$, $BF_{10} = 2.38$), suggesting that participants rapidly adjusted to new distractor contexts.

Second, we analysed whether repeating diagnostic dimensions across trials facilitated performance in an exploratory analysis for the interleaved condition. We computed the LISAS separately for trials in which the diagnostic dimension was the same as on trial n-1 and trials for which the diagnostic dimension had switched from the previous trial, only including those trials on which the previous response had been correct, but independent of the context cues. Indeed, performance tended to be slightly higher when the diagnostic dimension was repeated, although this was not significant ($t(79) = 1.89$, $p = .06$, $d_z = 0.21$, $BF_{10} = 0.66$).

To test whether some participants may still have successfully used the context cues, we also computed the correlation of context facilitation in the blocked and interleaved conditions across participants. This correlation was not significant ($\rho = -.07$, $p = .52$).

Besides the effects of context expectations (or their absence), performance was again always better on unique rectilinearity trials in both blocked and interleaved condition (main effects of trial type, *p*'s < .001).

Discussion

As in our previous experiments, we found that blocked distractor context modulated the attentional template, with better performance when the diagnostic dimension on a trial was expected rather than unexpected.

When distractor context was interleaved, we no longer found an effect of distractor expectations. While the attentional template was initially biased to emphasise diagnostic target dimensions in the blocked context, it was not influenced by context cues in the interleaved condition and therefore likely included both features of the shapes. Importantly, this was the case even though participants were explicitly informed about the contingencies, were first exposed to them in a blocked manner and reminded about them prior to every block.

Our results showed that explicit context cues were not readily exploited to solve the task, compared to the apparent ease with which regularities in target-distractor regularities were extracted and used when context was blocked. On the one hand, this could suggest that target-distractor relations are generally not learnt in a context-dependent manner. Distractor context effects would then reflect biases based on recent selection history, where target-distractor statistics of previous trials were integrated in the search template independent of context. Accruing over multiple trials, these biases may be stronger than observed in the interleaved condition.

On the other hand, it may still be possible to update the attentional template on a trial-by-trial basis under different conditions. Quickly adjusting to abstract, explicit context cues may be challenging, and not always done by participants, depending perhaps on the amount of training with those cues and whether they are necessary to do the task. In the orientation discrimination study mentioned earlier, participants underwent three experimental sessions (Scolari et al., 2012) and the task required very fine-grained discriminations. The same cues did not affect sensory gain when discrimination was easy (Scolari & Serences, 2009). In a similar vein, when participants searched for a specific watch among multiple instances of another watch, cueing the identity of the distractor watch in advance facilitated search (relative to when there was no cue), but only when participants had not been well familiarised with the target watch prior to search, which may have rendered the task easier (Bravo & Farid, 2016).

While the background colours used were salient, attending to them was not strictly necessary to find the target and they may therefore have been ignored. On the other hand, our results also suggested that context only needed to remain stable over relatively few trials for participants to adjust to context. To test whether it was still possible to emphasise different target dimensions on a trial-by-trial basis, we decided to provide stronger context cues, while keeping the overall task and difficulty equal.

Experiment 4 – Adjusting to distractor context on a trial-by-trial basis with location as context cue

In naturalistic search, different target-distractor relations are often found in different locations, and in addition to visual differences between contexts, spatial location is likely a very strong cue to separate contexts. Memory can strongly depend on spatial context (Godden & Baddeley, 1975; Smith et al., 1978; Smith & Vela, 2001), and spatial separation may therefore help to separate memories of previous searches in both contexts, which may otherwise interfere with each other. If that is the case, combining the different colour contexts with additional (redundant) spatial information, presenting them at different screen locations, may yield stronger context-dependent learning.

Therefore, in Experiment 4 we followed the same approach as in Experiment 3, again testing whether diagnostic target dimensions could be learnt in a context-dependent manner, but distractor contexts were also separated in space. Two coloured squares were continuously presented on the left and right of the screen, with participants' attention being cued towards one of the two contexts at the start of each trial.

Methods

Participants

Eighty participants (47 females, mean age: 28.48, sd: 4.23) were included in this new experiment. Additional participants were replaced due to not completing the experiment/having a response rate < 85% (N=4), because their overall performance was < 55% (N=7) or were at chance when a familiar target shape was used as distractor (N=4). Sample size was chosen for consistency with Experiment 3 and again allowed us to detect main effect of context match effect about half as large as the blocked condition in Experiment 1 with 80% power ($\eta_p^2 > .095$, with $\alpha = .05$).

All participants were recruited via Prolific (https://www.prolific.co), provided informed consent and received £5.00 for their participation. The study was approved by the Radboud University Faculty of Social Sciences Ethics Committee (ECSW-LT-2022-9-9-25489). All data were collected in 2022.

Procedure

The overall training and search task were similar to Experiment 3, including the blocked context blocks and the number of trials, with the exception that the two distractor contexts were separated in space (Figure 7). Two coloured squares (blue-green and magenta) were constantly presented on the left and right of the display, separated by a thin white line in the display centre. At the onset of a trial, the name of a shape appeared in one of the squares, designed to direct participants' fixation towards one context, followed by the search scene. Feedback was given by a coloured fixation dot appearing at the centre of the screen after each trial. The search display always appeared in the cued colour square (while the other one remained empty), but the redundant spatial and colour cue predicted the upcoming trial type with 80% validity.

Target orientation, rectilinearity and validity were counterbalanced per block, while the specific target individuals and the display side on which the target appeared were counterbalanced across the whole experiment. All factors were counterbalanced separately for both contexts. The distractor context participants started with, the associations between block context, background colour and display sides were all randomised.

Stimuli & Setup

Search displays were created as in Experiment 3, but their size on screen was reduced (to approximately 15 x 0.9° assuming a 57 cm viewing distance). The online experiment was programmed in Psychopy and hosted on Pavlovia (https://pavlovia.org).

Data Analysis

Distractor context effects were again analysed separately for the blocked and interleaved context blocks with two 2 (trial type: rectilinearity, unique orientation trial) x 2 (context match: matching, mismatching context cue) repeated measures ANOVAs on the LISAS.

In addition, we compared the interleaved conditions of Experiment 3 and 4, by using a 2 (trial type: rectilinearity, unique orientation trial) x 2 (context match: matching, mismatching context cue) x 2 (Experiment) repeated measures ANOVA, with experiment as between-subjects factor.

Context matching effects for individual dimensions were followed up with t-tests (two-sided, $\alpha = .05$). Bayes factors for simple t-contrasts and the main effects/ interactions were calculated as Bayesian one-sample t-tests for the respective difference scores between conditions or condition averages using the JASP default settings (Cauchy prior with scale 0.707). For comparisons between experiments or groups, Bayesian independent sample t-tests were computed instead.

Results

Overall accuracy and RT were 86.32% (sd 8.52) and 749.73 ms (sd 114.41), respectively. Following our preregistration and the procedure described for Experiment 3, anticipatory and delayed responses (on average 1.42% (sd 0.86) of trials) were excluded.

Again replicating our previous findings, context expectations influenced search in the blocked condition, with better performance when the diagnostic target dimensions matched participants' expectations (Figure 9; main effect of context match: $F(1,79) = 13.32$, $p <.001$, $\eta_p^2 = .14$, $BF_{10} = 47.98$). This was driven mostly by the unique rectilinearity trials (context match x trial type: $F(1,79) = 5.04$, $p = .03$, $BF_{10} = 1.31$; rectilinearity trials: $t(79) = 3.73$, $p < .001$, $d_z = 0.417$, $BF_{10} = 61.18$; orientation trials: $t(79) = 0.35$, $p = .725$, $d_z = 0.04$, $BF_{10} = 0.13$).

Most importantly, the attentional template was also shaped by context cues in the interleaved condition, resulting in better performance when the diagnostic target dimension was expected (context match: $F(1,79) = 7.98$, $p = .006$, $\eta_p^2 = .09$, $BF_{10} = 4.89$), with no difference between diagnostic dimensions (context match x trial type: $F(1,79) = 2.31$, $p = .13$, $\eta_p^2 = .02$, $BF_{10} = 0.37$; context match unique rectilinearity trials: $t(79) = 2.90$, p = .005, $d_z = 0.32$, $BF_{10} = 5.60$; orientation $t(79) = 1.02$,

$p = 0.311$, $d_z = 0.11$, $BF_{10} = 0.20$). Independent of context expectations, the target was always more easily found on rectilinearity unique trials in both the blocked and interleaved condition (p's < .001).

The effect of context expectations in the interleaved condition remained consistent when combining all data from Experiments 3 and 4 (context match: $F(1,158) = 8.80$, $p = .003$, $\eta_p^2 = .05$, $BF_{10} = 5.85$), and the effect did not significantly differ across experiments (context match x experiment: $F(1,158) = 0.22$, $\eta_p^2 = .01$, $BF_{10} = 0.34$).

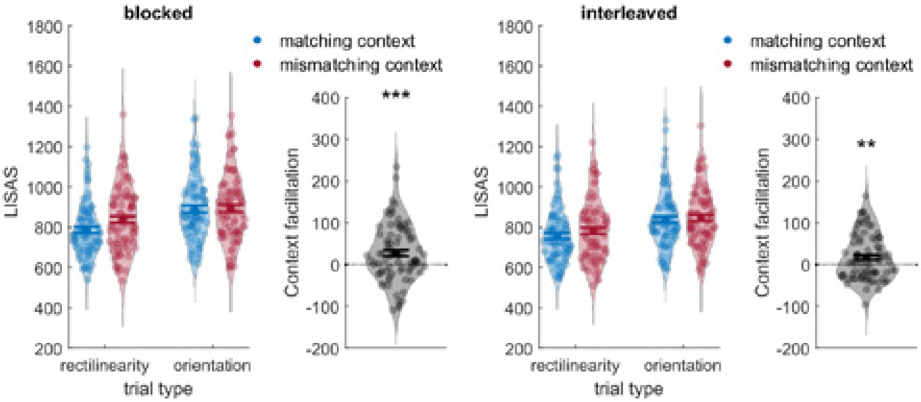

Figure 9: Results of Experiment 4. Performance on both trial types as a function of context (left subpanel) and context facilitation (LISAS mismatching—matching context) averaged across trial types (right subpanel). All error bars are SEM. LISAS = linear integrated speed accuracy score; SEM = standard error mean.

Given that participants formed context expectations in both the blocked and interleaved condition, we again tested whether context facilitation in both conditions correlated across participants, to complement our main preregistered analysis. This was however not significant, neither in Experiment 4 alone ($\rho = 0.16$, $p = 0.19$), nor in Experiments 3 and 4 combined ($\rho = .06$, $p = .47$).

Discussion

As in Experiment 3, we tested whether target-distractor relations could be learnt in a context-dependent manner and bias the attentional template on a trial-by-trial basis, but using spatial location as additional context cue. Spatial location may support context-depending learning, although a direct comparison between both experiments did not show significantly stronger context facilitation.

Most importantly, we now found that context cues modulated the attentional template on a trial-by-trial basis, indicating that the target-distractor relations were learnt in a contextualised manner and that the attentional template was biased by long term memory for the target-distractor relations in different contexts. Thus, distractor context does not only modulate the attentional template based on context-independent short-term priming mechanisms (including those that may extend over multiple trials), but can also flexibly bias the template, top-down, whenever a familiar distractor context is re-encountered.

Overall, we did not observe strong correlations between context facilitations in blocked and interleaved conditions, even across a large sample (N=160). In stable and rapidly varying distractor contexts, adjusting the attentional template to distractor expectations may not rely on fully overlapping processes. One possibility is that context-independent biases, based on selection history, additionally contribute to context facilitation in the blocked condition but not the interleaved condition.

General discussion

Our visual system selects those objects and features for further processing on the basis of their match with an internal representation of our current goals (J. Duncan & Humphreys, 1989; Eimer, 2014; Wolfe, 2021). Which visual features allow for efficient search, however, depends not only on the visual features of the target, but also on how they relate to the features of the current distractors (J. Duncan & Humphreys, 1989; Geng & Witkowski, 2019; Navalpakkam & Itti, 2007). To test the flexibility of adaptive biases in attentional templates based on distractor expectations, here we investigated how distractor expectations shape the target template for complex 2D shapes in both blocked and interleaved distractor contexts. Participants searched for these targets in two probabilistic contexts, rendering either the target's rectilinearity or orientation diagnostic and we tested whether target features in the expected diagnostic dimension were emphasised relative to the non-diagnostic one.

Across multiple experiments, we found that participants learnt the statistical regularities of different distractor contexts when these were blocked, such that search performance was better on trials matching the most frequent diagnostic dimension of a block. The same was found for search trials with only one distractor (Experiment 2), confirming that the attentional template for the target was biased by distractor expectations in a way that emphasised the most diagnostic target dimension. This bias was, however, relative rather than absolute. That is, participants did not simply look for, e.g., any horizontal shape entirely independent of its rectilinearity, as performance on mismatching trials and all two-shape trials (Experiment 2) was still much higher than chance.

In addition, our results suggest adjusting to distractor context may also be possible without participants noticing these regularities. Over the course of short blocks, participants extracted regularities about common diagnostic feature dimensions from the search scene, independent of specific target or distractor features. Context facilitation was equal for participants who reported being aware of the context manipulation and those who did not (Experiment 2). It is less clear whether the same is true for the interleaved condition in Experiments 3 and 4. The mapping between context cues and target-distractor relations may also be learnable implicitly, but we cannot rule out explicit knowledge was necessary in our task. Nonetheless, this knowledge by itself was not sufficient to use context cues in Experiment 3. Further, in our daily life environments, where contexts are highly familiar and rich in cues to separate them, adjusting attentional target templates in a context-dependent manner may be rather effortless.

Finally, in Experiment 4, but not Experiment 3, we found that predictive distractor context was also exploited when context changed on a trial-by-trial basis, showing that distractor context also shaped the template independently of biases building up over immediately repeating searches (i.e. selection history). This is relevant as it demonstrates that attentional templates can be truly context-dependent.

Together, these findings provide important evidence that even attentional templates for complex shapes, retrieved from memory, can be flexible to optimally support search performance by highlighting diagnostic information. That being said, it is unlikely that searchers rely on entirely distinct and non-overlapping attentional templates across contexts, as those templates may still be based on the same underlying memory representation (Boettcher et al., 2020; X. Yu, Zhou, et al., 2023) and within most contexts there will always be some uncertainty about the diagnosticity of certain target features before the object is actually identified. As mentioned already, our results suggest a relative bias in how strongly a given target feature is emphasised in the template, depending on distractor expectations, rather than a total change of the template; accordingly, our effect sizes are relatively small. Whether context cues are used and the template adjusted likely depends on whether context-dependent templates are necessary for successful search (Bravo

& Farid, 2016; Scolari et al., 2012; X. Yu, Zhou, et al., 2023) and how easily separate target representations in different contexts are formed and retrieved from memory (where space could act as a relevant additional context cue). Indeed, the null result we observed for the interleaved condition in Experiment 3 suggests there are also clear limits to the flexibility of the template.

Importantly however, we found that attentional templates were surprisingly flexible under conditions that could have promoted unbiased templates instead, for several reasons. First, targets and distractors were well distinguishable and participants were familiarised with the target shapes outside of a search context, which may promote a distractor-independent representation (cf. Bravo & Farid, 2016). Second, the two dimensions together formed the outline shape of the targets, such that the dimensions were not clearly separable. Third, the use of symbolic cues prevented the selective encoding of one dimension of the current visual input over another (Boettcher et al., 2020), as could be the case when using picture cues. Finally, attentional templates were biased by distractor expectations even when distractor context changed frequently, on a trial-by-trial basis. This indicates that biases in the template can indeed arise relatively efficiently and effortlessly, making it likely that similar mechanisms operate in daily-life visual search.

These findings were obtained from a larger sample of online participants, outside of a typical lab context and are supported by prior studies using similar search tasks with simpler stimuli and a different combination of dimensions (Boettcher et al., 2020; J. Lee & Geng, 2020). While this provides evidence that context expectations can influence different dimensions of the target template, including rectilinearity as an important dimension for real-world shapes, further dimensions (e.g., size or ‘spikyness’) remain to be tested. Furthermore, it is still unclear whether context facilitation is more strongly observed for target dimensions that are overall prioritised in the template, as suggested by our results in Exp 2A and 3. Indeed, we found differences in context facilitation between target dimensions in some experiments, but these were not consistently present and the two dimensions were not carefully matched in how well they generally allowed to distinguish targets from distractors; it is well possible that orientation is more strongly modulated by context expectations in other settings. More generally, whether distractor expectations bias the template may depend on the relative advantage of including these expectations compared to a stable template (based on the predictability of specific distractor and target features within and across different contexts). Difficult search tasks with few, highly predictable distractor contexts and dimensions that can be independently represented by the visual system may lead to stronger context effects than searches across many variable and less predictable contexts.

We interpret our results in terms of biased attentional templates rather than, for example, improvements in distractor grouping (Experiment 2). However, independent of grouping, in principle it is possible that expected distractor features were suppressed better (see van Moorselaar & Slagter, 2020 for a review) rather than target features being enhanced. We think this is unlikely, because cueing distractor features on a trial-by-trial basis often leads to no benefits, or even costs (e.g., M. W. Becker et al., 2016; Cunningham & Egeth, 2016). Furthermore, while feature-based inhibition based on statistical learning and blocked distractors is possible (Stilwell et al., 2019; van Moorselaar et al., 2020, 2021; Vatterott & Vecera, 2012), it is likely not context-dependent (see Britton & Anderson, 2019; de Waard et al., 2022 for space-based suppression). Successful feature-based distractor inhibition in our experiments would require learning specific distractor features associated with every target in both contexts. Enhancement of diagnostic target features therefore seems a more probable mechanism.

Where and when does this enhancement take place? A relevant direction for future research will be to establish whether distractor context expectations influence early attentional selection and guidance, or alternatively only later decision processes. Attentional selection and target-identification might rely on different templates (Wolfe, 2021) and those may be differentially shaped by distractor context (Hamblin-Frohman & Becker, 2021; X. Yu, Hanks, et al., 2022). At the neural level, attentional templates have been associated with a baseline increase of neural activity in neurons tuned to target features (Desimone & Duncan, 1995), which serves to bias competition in favour of the target in a subsequently presented search array (see e.g. Battistoni et al., 2017; Eimer, 2014 for reviews). A bias in the template could thus be instantiated by an increase in preparatory activity of neurons coding for the discriminative feature of the target and lead to an increased response of these neurons when the target is presented. In line with this, Scolari et al. (2012) found that the gain of V1 neurons whose activity discriminated best between target and distractors in a difficult orientation discrimination task was selectively enhanced. In the current study, looking for, e.g., a rectilinear shape in rectilinearity unique trials could have increased baseline activity or gain of neurons preferring rectilinear features. Context-dependent preparatory biases may influence early sensory processes, but additional neuroimaging studies are needed to understand where in the visual hierarchy and at which processing stages context expectations influence search.

During real-world search, the attentional template may be informed by context also in other ways. For example, when experiencing the same object (or object category) across multiple contexts, we may learn which features are diagnostic of this object across different contexts and selectively include those diagnostic parts in the template (Reeder & Peelen, 2013). Which features are diagnostic may not only reflect the relation of targets and distractors, but also the variability of the target itself (i.e., which features consistently characterise exemplars of a category or the same object across time). Along these lines, there is evidence that target-dimensions with lower variance are emphasised in the template for target-match decisions (Witkowski & Geng, 2019, 2022). This can be seen as long-term biases based on extensive experience. While there may therefore be less need to adjust these object representations to respective contexts, they are still not static and can also be biased by current context and expectations. Predictive distractor contexts are only one instance of regularities shaping the attentional template. For example, recent studies found that participants are sensitive to colour statistics within specific object categories (Bahle et al., 2021) also in a context-dependent manner (Kershner & Hollingworth, 2022). In addition, relations between objects and the overall scene context (e.g., the expected retinal size of an object at a particular location) are encoded in the attentional template (Gayet & Peelen, 2022), all providing evidence for the adaptability of attentional templates.

To conclude, we provide evidence for flexible and adaptive biases in the attentional template when searching for the same objects within different contexts. Distractor context was learnt quickly and seemed to be incorporated in the template independently of awareness of the context manipulation. Moreover, attentional templates were biased by distractor expectations in a context-dependent manner, even when context was interleaved. Further investigating how and when context is used to shape attentional allocation will help to understand how we efficiently find objects in our daily-life environments.

Supplementary materials

Results Accuracy & RT – Experiment 1

We conducted the same 2 (trial type: unique rectilinearity, unique orientation trial) x 2 (context match: presented in matching, mismatching context block) repeated-measures ANOVA reported for the LISAS in the main text also separately for accuracy and RT.

For accuracy, the main effect of context match was significant ($F(1,33) = 8.05$, $p = .008$, $\eta_p^2 = .20$, $BF_{10} = 5.37$). Averaged over trial types, context facilitation was 3.54% (sd 7.16). In addition, there was a significant main effect of trial type, with better performance on rectilinearity unique trials ($F(1,33) = 8.53$, $p = .006$, $\eta_p^2 = .21$, $BF_{10} = 6.45$). Both factors did not interact ($F(1,33) = 0.2$, $p = .65$, $\eta_p^2 = .01$, $BF_{10} = 0.20$; context match unique rectilinearity trials alone ($t(33) = 1.68$, $p = .10$, $d_z = 0.29$, $BF_{10} = 0.65$); unique orientation trials ($t(33) = 2.32$, $p = .03$, $d_z = 0.29$, $BF_{10} = 1.89$).

For reaction times, context facilitation was 7.50 ms (sd 34.64), but the main effect of context match not significant ($F(1,33) = 1.59$, $p = .22$, $\eta_p^2 = .05$, $BF_{10} = 0.38$). There was also a significant main effect of trial type ($F(1,33) = 78.52$, $p < .001$, $\eta_p^2 = .70$, $BF_{10} > 3 \times 10^7$). There was no significant interaction between trial type and context match ($F(1,33) = 0.14$, $p = .71$, $\eta_p^2 = .004$, $BF_{10} = 0.20$; unique rectilinearity: $t(33) = 1.00$, $p = .34$, $d_z = 0.17$, $BF_{10} = 0.29$, unique orientation: $t(33) = 0.47$, $p = .64$, $d_z = 0.08$, $BF_{10} = 0.20$).

Results Accuracy & RT – Experiment 2

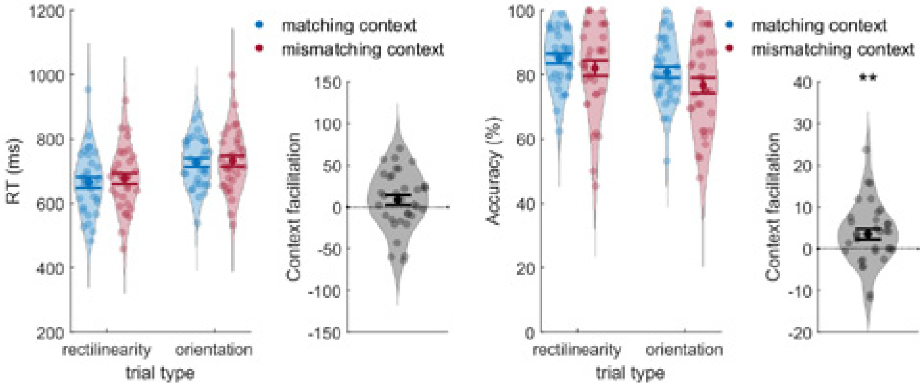

Figure 10: Results of Experiment 1. Performance (accuracy and RT) on both trial types as a function of block context (left subpanel) and context facilitation (difference between matching and mismatching) averaged across trial types (right). All error bars are SEM.

Experiment 2A

For accuracy, there was no significant main effect of context match ($F(1,33) = 0.3$, $p = .59$, $\eta_p^2 = .01$, $BF_{10} = 0.21$). The main effect of trial type was significant ($F(1,33) = 5.85$, $p = .02$, $\eta_p^2 = .15$, $BF_{10} = 2.21$), due to better performance on rectilinearity unique trials. Both factors did not interact ($F(1,33) = 0.09$, $p = .76$, $\eta_p^2 = .00$, $BF_{10} = 0.19$; context match unique rectilinearity trials: $t(33) = -0.69$, $d_z = -0.12$, $p = .50$, $BF_{10} = 0.23$; orientation trials: $\mathrm{t}(33) = -0.12$, $d_z = -0.02$, $p = .91$, $BF_{10} = 0.19$).

For reaction times, we found a significant main effect of match ($F(1,33) = 4.65$, $p = .038$, $\eta_p^2 = .12$, $BF_{10} = 1.41$), reflecting a context facilitation of 10.87 ms (sd 29.36) across trial types. There was also a significant main effect of trial type ($F(1,33) = 46.43$, $p < .001$, $\eta_p^2 = .58$, $BF_{10} > 1\mathrm{x}10^7$) and a significant interaction between both ($F(1,33) = 8.01$, $p = .008$, $\eta_p^2 = .20$, $BF_{10} = 5.30$). This reflected significant context facilitation for the rectilinearity trials alone ($t(1,33)$, 3.61, $p = .001$, $d_z = 0.62$, $BF_{10} = 32.10$) but no facilitation in the orientation trials ($t(1,33) = -1.26$, $p = .21$, $d_z = - 0.21$, $BF_{10} = 0.38$).

Experiment 2B

In accuracy, the main effect of context match was not significant ($F(1,79) = 1.53$, $p = .21$, $\eta_p^2 = .019$, $BF_{10} = 0.26$). As performance was better on rectilinearity unique trials, the main effect of trial type was significant ($F(1,79) = 16.77$, $p < .001$, $\eta_p^2 = .175$, $BF_{10} = 196.68$). There was no significant interaction between both ($F(1,33) = 1.53$, $p = .21$, $\eta_p^2 = .004$, $BF_{10} = 0.15$; context match rectilinearity trials: $t(79) = 0.33$, $p = .74$, $d_z = 0.03$, $BF_{10} = 0.13$; orientation trials: $t(79) = 1.21$, $p = .29$, $d_z = 0.14$, $BF_{10} = 0.25$).

For RT, the main effect of context match was significant ($F(1,79) = 13.38$, $p < .001$, $\eta_p^2 = .145$, $BF_{10} = 28.46$), reflecting a context facilitation of 10.89 ms (sd 26.61). We also found a significant main effect of trial type ($F(1,79) = 137.98$, $p < .001$, $\eta_p^2 = .63$, $BF_{10} > 5 \times 10^{15}$). Trial type and context match did not interact significantly ($F(1,79) = 0.11$, $p = .74$, $\eta_p^2 = .00$, $BF_{10} = 0.13$; context match rectilinearity trials: $t(79) = 1.60$, $p = .11$, $d_z = 0.21$, $d_z = 0.18$, $BF_{10} = 0.42$; orientation trials $t(79) = 1.92$ $p = .06$, $d_z = 0.21$, $BF_{10} = 0.71$).

Comparing context facilitation for participants who had high or no awareness of the context manipulation combined for all participants of Experiment 2A and 2B, context facilitation for RT was not significant for the no-awareness group ($F(1,34) = 0.83$, $p = .37$, $\eta_p^2 = .25$, $BF_{10} = 0.27$), but was found for the high-awareness group ($F(1,33) = 10.76$, $p = .003$, $\eta_p^2 = .25$, $BF_{10} = 14.51$) and the difference between both did not reach significance ($t(67) = -1.93$, $p = .06$, $d = -0.46$, $BF_{10} = 1.18$).

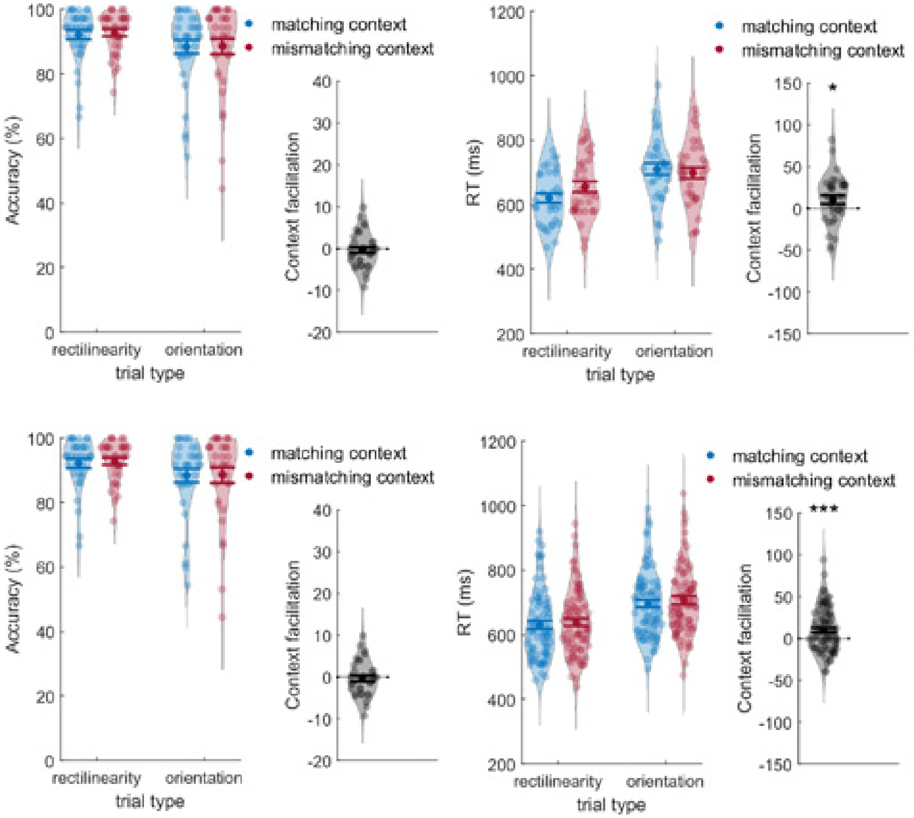

Figure 11: Results of Experiment 2. Performance (accuracy and RT) on both trial types as a function of block context (left subpanel) and context facilitation (difference between matching and mismatching) averaged across trial types (right). Upper panels show results for Experiment 2A, lower for Experiment 2B, All error bars are SEM.

Results Accuracy & RT – Experiment 3

In the blocked context blocks, we observed a main effect of context match ($F(1,79) = 5.69$, $\eta_p^2 = 0.067$, $p = 0.02$, $BF_{10} = 1.76$) in accuracy, with context facilitation of 2.11% (sd 7.91) averaged over trial types. There was also a main effect of trial type, due to better performance for rectilinearity unique trials ($F(1,79) = 25.32$, $p < .001$, $\eta_p^2 = .248$, $BF_{10} = 147.00$). The interaction between trial type and context match was not significant ($F(1,79) = 1.84$, $p = .18$, $\eta_p^2 = .022$, $BF_{10} = 0.30$; context match: unique rectilinearity trials: $t(79) = 2.94$, $p < 0.01$; unique orientation: $t(79) = 0.46$, $p = 0.65$; context match unique orientation trials ($t(79) = 2.94$, p = 0.004, $d_z = 0.33$, $BF_{10} = 6.57$, unique orientation trials t(79) = 0.46, p = 0.647, $d_z = 0.05$, $BF_{10} = 0.13$).

For RT, there was also a main effect of context match, context ($F(1,79) = 6.81$, $p = .01$, $\eta_p^2 = .079$, $BF_{10} = 2.68$) reflecting a 12.60 ms (sd 43.16) context advantage averaged over trial types. There was also a main effect of trial type type ($F(1,79) = 82.507$, $p < .0001$, $\eta_p^2 = .510$, $BF_{10} > 4000$), and a significant interaction between both ($F(1,79) = 6.818$, $p = .001$, $\eta_p^2 = .123$, $BF_{10} = 24.30$). This reflected a significant match effect for rectilinearity alone ($t(1,79) = 4.36$, $p < .0001$, $d_z = 0.487$, $BF_{10} = 438.53$), but a numerically negative context facilitation for orientation ($t(1,79) = - 1.58$, $p = .11$, $d_z = 0.18$, $BF_{10} = 0.41$)

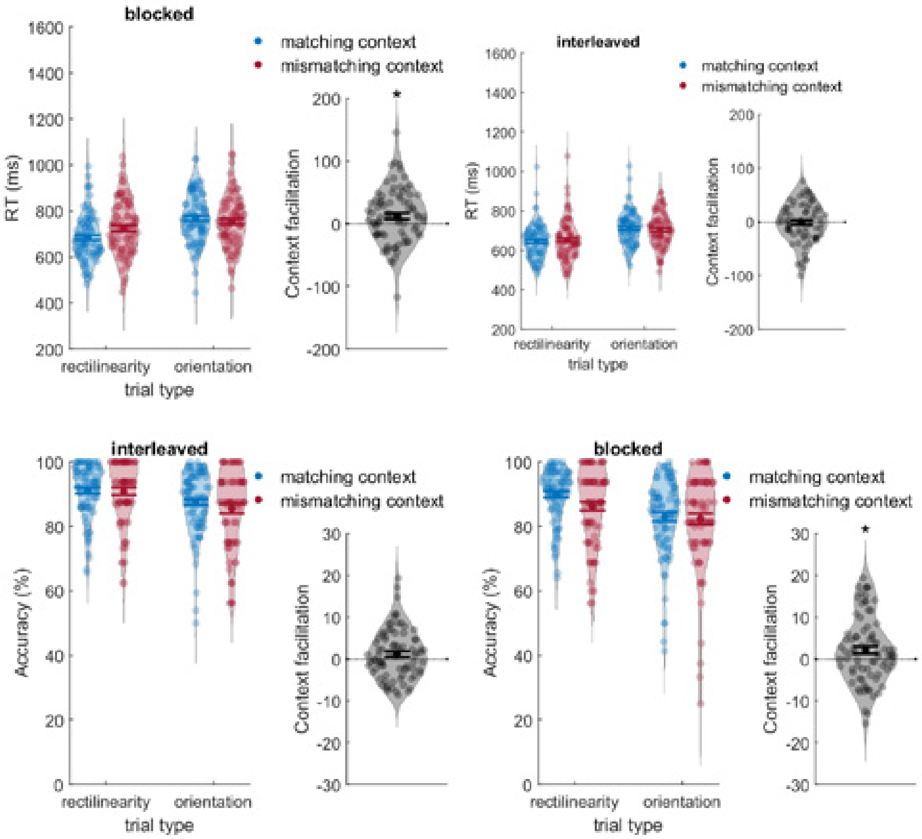

Figure 12: Results of Experiment 3. Performance (accuracy and RT) on both trial types as a function of context (left subpanel) and context facilitation (difference between matching and mismatching contexts) averaged across trial types (right subpanel). All error bars are SEM.

In the interleaved condition, the main effect of context match did not reach significance for accuracy ($F(1,79) = 3.11$, $p = .08$, $\eta_p^2 = .038$, $BF_{10} = 0.54$), reflecting a 1.17 % (sd 5.91) context advantage. The main effect of trial type was again significant ($F(1,79) = 29.02$, $p < .001$, $\eta_p^2 = 0.269$, $BF_{10} = 86.04$) and there was no interaction between both (F(1,79) = 2.37, p = .12, η_p^2 = .029, BF_{10} = 0.38; context match unique rectilinearity: t(79) = 0.32, p = 0.75, d_z = 0.04, BF_{10} = 0.13; unique orientation: t(79) = 2.16, p = 0.03, d_z = 0.24, BF_{10} = 1.11). There was no significant effect of the diagnostic dimension of the previous trial on accuracy in the current trial (t(79) = 1.23, p = 0.22, d_z = 0.14, BF_{10} = 0.25).

For RT, the main effect of match was not significant (F(1,79) = 0.08, p = .76, η_p^2 = .001, BF_{10} = 0.13). There was, however, again a main effect of trial type (F(1,33) = 107.3, p < .001, η_p^2 = .58, BF_{10} > 1 x 10^6) and no significant interaction between both factors (F(1,79) = 2.65, p = .11, η_p^2 = .032, BF_{10} = 0.43; context match unique rectilinearity trials: t(79) = 1.10, p = 0.273, d_z = 0.12, BF_{10} = 0.22 ; unique orientation: t(79) = -1.32, p = 0.19, d_z = - 0.15, BF_{10} = 0.28). The diagnostic dimension of the previous trial did also not significantly affect RT on the current trial (t(79) = 1.00, p = 0.32, d_z = 0.11, BF_{10} = 0.20).

Context facilitation in both conditions did not correlate significantly (accuracy: ρ = .02, p = .86; RT: ρ = .03, p = .78).

Results Accuracy & RT – Experiment 4

In the blocked condition, the main effect of context match for accuracy was significant (F(1,79) = 11.50, p = 0.001, ηp^2 = 0.13, BF_{10} = 22.43), reflecting an overall advantage of 3.37 % (sd 8.90) for trials in matching contexts. We also found a main effect of trial type (F(1,79) = 32.35, p < .001, ηp^2 = 0.29, BF_{10} = 602.43) and no interaction between both (F(1,79) =1.11, p = .27, η_p^2 = .01, BF_{10} = 0.21; context match rectilinearity trials: t(79) = 3.09, p = 0.003, d_z = 0.35, BF_{10} = 9.73; orientation: t(79) = 1.62, p = 0.11, d_z = 0.18, BF_{10} = 0.43).

For RT the main effect of context match failed to reach significance (F(1,79) = 3.53, p = .06, η_p^2 = .04, BF_{10} = 0.67) and the main effect of trial type again significant (F(1,79) = 74.00, p < .001, η_p^2 = .48, BF_{10} = 19.62). Both factors interacted (F(1,79) = 5.27, p = .02, η_p^2 = .06, BF_{10} = 1.46), reflecting significant context facilitation for unique rectilinearity trials (t(79) = 3.08, p = .003, d = 0.34, BF_{10} = 9.54), but not unique orientation trials (t(79) = -0.86, p = .39, d_z = -0.10, BF_{10} = 0.18).

In the interleaved condition, we also found a significant effect of context match on accuracy (F(1,79) = 7.41, p = .008, η_p^2 = .09, BF_{10} = 3.81), with overall context facilitation of 2.09% (sd 6.85). There was also a significant main effect of trial type (F(1,79) = 28.18, p < .001, ηp^2 = .26, BF_{10} = 37.65), but no interaction between both (F(1,79) = 0.48, p = .49, η_p^2 = .01, BF_{10} = 0.16; context match unique rectilinearity:

$t(79) = 1.51$, $p = 0.13$, $d_z = 0.24$, $BF_{10} = 0.36$; orientation: $t(79) = 2.17$, $\mathrm{p} = 0.03$, $d_z = 0.24$, $d_z = 0.17$, $\mathrm{BF}_{10} = 1.13$).

For RT, the overall effect of context match was not significant ($F(1,79) = 1.17$, $p = .28$, $\eta_p^2 = .01$, $BF_{10} = 0.22$). There was, however, again a main effect of trial type ($F(1,79) = 74.94$, $p < .0001$, $\eta_p^2 = .49$, $BF_{10} = 1.02$) and a significant interaction between trial type and context match type ($F(1,79) = 4.49$, $p = .04$, $\eta_p^2 = .05$, $BF_{10} = 0.02$). Context facilitation was significant for unique rectilinearity trials ($t(79) = 2.39$, $p = .02$, $d_z = 0.27$, $\mathrm{BF}_{10} = 1.78$), but not unique orientation trials ($t(79) = -0.88$, $p = .38$, $d_z = -0.10$, $\mathrm{BF}_{10} = 0.18$).

Context facilitation in both conditions did not correlate significantly (accuracy: $\rho = .18$, $p = 0.18$; RT: $\rho = .14$, $p = 0.21$).

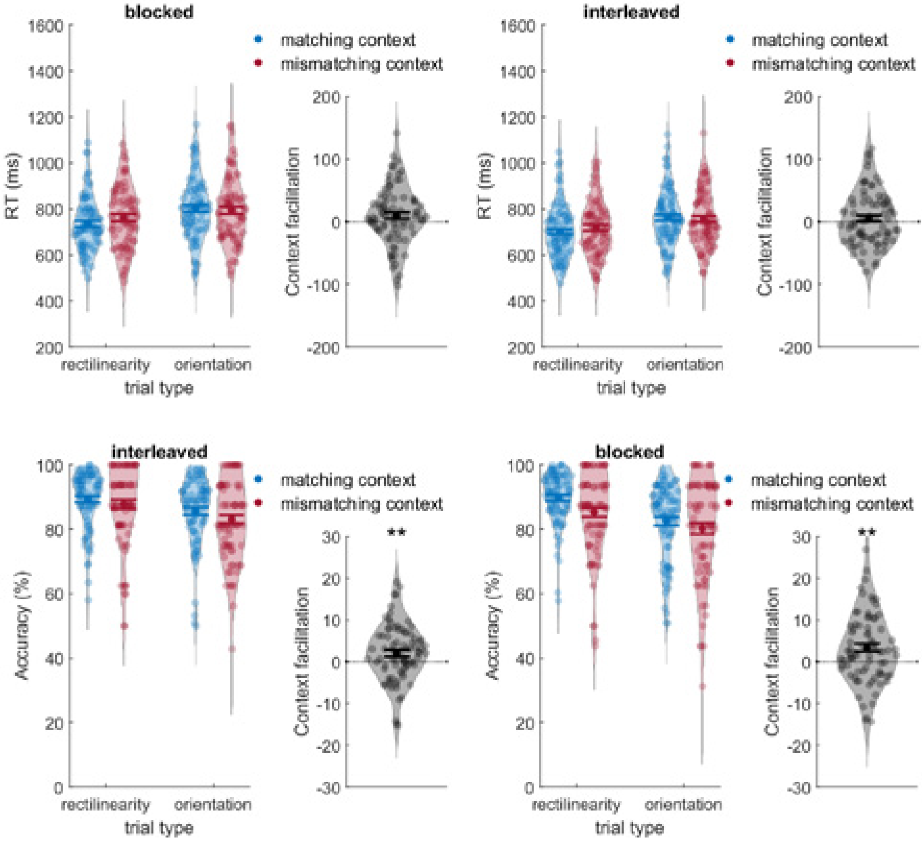

Figure 13: Results of Experiment 4. Performance (accuracy and RT) on both trial types as a function of context (left subpanel) and context facilitation (difference between matching and mismatching contexts) averaged across trial types (right subpanel). All error bars are SEM.

Chapter 3

Attentional Guidance Through Object Associations in Visual Cortex

Efficient behaviour requires the rapid attentional selection of task-relevant objects. Preparatory activity of target-selective neurons in visual cortex is thought to support attentional selection, guiding spatial attention and favouring processing of target-matching input. However, naturalistic searches are often guided by non-targets, including target-associated "anchor" objects. For instance, when looking for a pen, we may direct our attention to the office desk on which we expect to find it. Here, using fMRI and eyetracking in a context-guided search task, we tested whether preparatory activity in visual cortex reflected the target, the guiding anchor object, or both. Participants learnt associations between targets and anchors, reversing across two scene contexts, before searching for these targets. Participants' first fixations were reliably guided by the associated anchor. Preparatory activity in lateral occipital cortex (LOC), and right intraparietal sulcus (IPS), represented the target-associated anchor rather than the target. These results shed light on the neural basis of context-guided search in structured environments.

This chapter is adapted from:
Lerebourg, M., de Lange, F. P., & Peelen, M. V. (2024). Attentional Guidance Through Object Associations in Visual Cortex. *Science Advances*.

Introduction

Imagine yourself standing in your colleague's messy office, looking for a pen to write down a note. Visual search tasks like these are challenging, yet we perform them frequently, efficiently, and seemingly without much effort. A seminal finding in the field of visual search is the finding that attentional selection of a target object is supported by preparatory activity of target feature-selective neurons in visual cortex (Desimone & Duncan, 1995), with preparatory activity targeting higher-level object representations in object-selective regions of the ventral stream when searching for objects in naturalistic scenes (Battistoni et al., 2017). For example, when monkeys were cued to search for a specific object, the firing rate of neurons in monkey inferotemporal cortex (IT) tuned to this object increased during the interval between the cue and the search display (Chelazzi et al., 1993, 1998). Similarly, in humans, fMRI activity patterns in visual cortex, especially object-selective regions within the ventral visual stream, represented the cued target in the absence of visual stimulation. This preparatory activity has been interpreted as an "attentional template" – a top-down bias in favour of target-matching input, resolving the competition between multiple visual stimuli, guiding attention and eye movements, and enhancing processing of potential targets (Bichot et al., 2005; Desimone, 1998; Eimer, 2014; Wolfe, 1994).

However, search in daily life differs from search in previously employed laboratory tasks in important ways. For example, in structured real-world scenes, search is not only guided by target features but also by contextually associated (non-target) objects (Boettcher et al., 2018; Koehler & Eckstein, 2017a, 2017b; Mack & Eckstein, 2011; Zhou & Geng, 2024). Going back to our example of looking for a pen, we may direct our attention to the office desk on which we expect the pen to be. In this case, the office desk reflects a so-called 'anchor' object (Boettcher et al., 2018; Draschkow & Võ, 2017; Helbing et al., 2022): it is large, salient (i.e., easy to find), associated with the pen, and provides spatial predictions about the pen's location. Whether or how preparatory activity can support context-guided search has not been previously investigated. Recent theories of visual search (Wolfe, 2021; X. Yu, Zhou, et al., 2023) distinguish between 'guiding templates' and 'target templates', but their respective link to preparatory activity is not yet clear (Dodwell & Eimer, 2024). We reasoned that if preparatory activity in visual cortex serves as an attentional guidance mechanism, it should not necessarily reflect the target, but rather those features or objects that are effective for guidance. If, instead, preparatory activity primarily supports target identification and/or processes supporting target-related decision making, it should reflect features of the target.

To test this, we designed a context-guided search task that allows for separating preparatory fMRI activity related to the target and to the anchor. Participants in the fMRI experiment were first familiarised with novel target-anchor associations, learning to associate two target categories (books and bowls) with two different tables across two rooms (e.g., the book was found on Table 1 in Room 1 but on Table 2 in Room 2; Figure 1). This design allowed us to test whether relevant guiding objects in the current context were represented in preparatory activity. The tables in our scenes were useful for search in a similar way as real-world anchors are, being easy to find and providing spatial predictions constraining the search space for the target, while controlling for visual or semantic similarity. After familiarisation, participants searched for the peripherally presented target objects. To measure preparatory activity for the target and anchor objects, we introduced preview-only trials, in which participants prepared to search for a target object within a given room, but the target and anchor objects did not appear in that trial. These trials allowed us to isolate fMRI activity patterns related to search preparation, and test whether those carried information about the associated anchor or, alternatively, only the current target.

Following previous work (Gayet & Peelen, 2022; Peelen & Kastner, 2011), our analyses focused on two visual cortex regions that may encode preparatory attentional templates: object-selective lateral occipital cortex (LOC) in the ventral visual stream and early visual cortex (EVC). Specifically LOC has most consistently been implicated in visual search for real-world targets (Gayet & Peelen, 2022; Peelen & Kastner, 2011; Soon et al., 2013; van Loon et al., 2018) and may therefore also represent a guiding template for anchor objects.

We found that the anchors guided eye movements, with the first fixation in the scene already directed towards the target-associated anchor, even in target-absent trials. Crucially, we found that preparatory fMRI activity in LOC during search preparation (preview-only trials), represented the target-associated anchor, independent of the target object and independent of the room that participants were searching. These results demonstrate that preparatory activity in visual cortex supports context-guided search by reflecting features relevant for guidance.

Results

Participants (N=34) searched for target objects (books and bowls) within two scene contexts (3D rendered images of a blue and yellow living room, see Figure 1) while undergoing fMRI scanning. The target objects were small and appeared in the periphery, placed among other non-target objects on one of two tables in the room. Within each room, targets of a given category always appeared on the same table, and the tables therefore acted as anchor-objects for the targets (100% validity). Across rooms, the target-anchor associations were switched (e.g., blue living room: book on Table 1, bowl on Table 2; yellow living room: book on Table 2, bowl on Table 1; specific associations counterbalanced across participants). Participants were familiarised with these associations before fMRI scanning (see Anchor-target association training). On each trial, a word cue ('book' or 'bowl') indicated the upcoming search target (1s), followed by a preview of the blue or yellow living room (3.1 – 4s), in which the tables were still occluded by grey rectangles. Together with the target cue, this preview could be used to prepare for the specific anchor-table associated with the target in the current room. Critically, within each room, the same table could appear either left or right with equal probability, and the preview was therefore not spatially predictive. All trial types (e.g., trials with different targets and rooms) were randomly intermixed, therefore requiring participants to retrieve the relevant target-anchor associations anew on each trial. On half of the trials (*search trials*), the occluders were then removed, briefly revealing the anchor-tables and objects placed on them. On all search trials, four different objects were placed on each table. On target absent trials, the target was substituted by an additional distractor object (e.g., a laptop, a coffee mug, or playing cards).

The search target was present on 50% of those trials. During this search phase, participants could freely inspect the scene and move their eyes. Average accuracy on those search trials was 72.46% (see Supplementary Materials for behavioural analyses). This shows that the search task was indeed challenging, making a preparatory template and contextual information useful for the task. Afterwards, a response screen appeared, prompting participants to report whether the target had been present or absent. On the other half of trials (*preview-only trials*), however, the occluders remained, and thus no objects or associated tables were shown. On those trials, participants were asked to keep fixating at the image centre and no response was required. Importantly however, since it was unknown during the preview whether the search scene would appear or not, participants still had to prepare to search on every trial. The preview-only trials were the trials of interest for the fMRI decoding, allowing us to isolate preparatory fMRI activity.

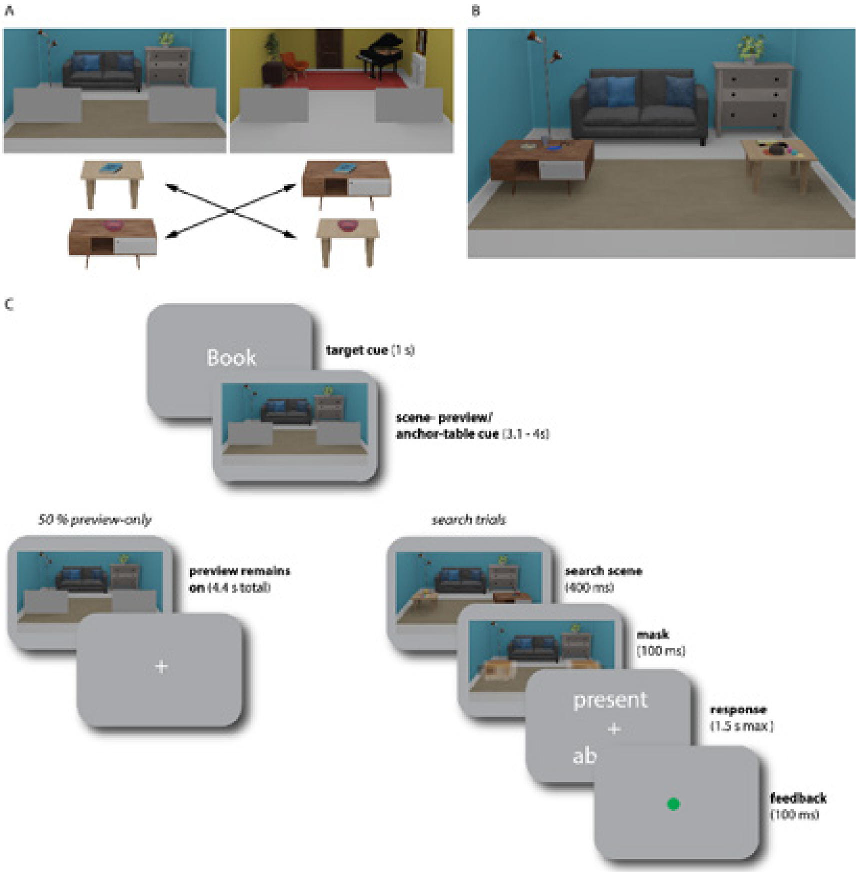

Figure 1: Experimental design and timeline. A Participants learnt room-dependent associations between target objects (books and bowls) and two different tables serving as anchor objects. **B** Example of a search scene. Both tables could either appear left or right in the scene, with multiple objects placed on them. If the associated target was present, it would appear on the associated table. Across the two rooms, associations between targets and anchors were switched. **C** Timeline of a trial. On each trial, participants were cued to search for one target category, followed by a scene preview. In search trials, a search scene would then briefly appear and participants were free to move their eyes. At the end of the trial, they indicated whether the cued target had been present. In half of the trials, however, participants prepared to search, but no search scene appeared after the preview and no response was required.

First fixations are guided by anchor objects

To test whether anchor objects guided attention, we analysed eye movements during the search phase, focusing on the first fixation in the scene (mean onset: 229.53 ms (sd 29.20)) as an index of overt attentional guidance. Target absent trials, in which target features could not guide fixations, provided a pure measure of guidance by the anchor. On those trials, 19.20% (sd 19.44) more first fixations were directed towards the correct anchor compared to the other non-associated anchor (CI = [12.50, 24.70], p < .001; Figure 2A). On target present trials, 25.24% (sd 20.59) more fixations were directed towards the correct anchor than the other non-associated anchor (CI = [19.00, 32.47], p < 0.001; Figure 2A), which was significantly higher than for target absent trials (CI = [0.40, 11.41], p = .03). This indicates that anchor features indeed guided attention, with additional guidance provided by target features when they were present. Finally, guidance by anchors also yielded search benefits, as indicated by a significantly positive correlation between anchor guidance and d': participants with more selective (anchor-guided) first eye movements on target absent trials also had a higher d' (r = .55, p < .001; Figure 2B). This correlation was also significant for target present trials (r= .60, p < .001; Figure 2B).

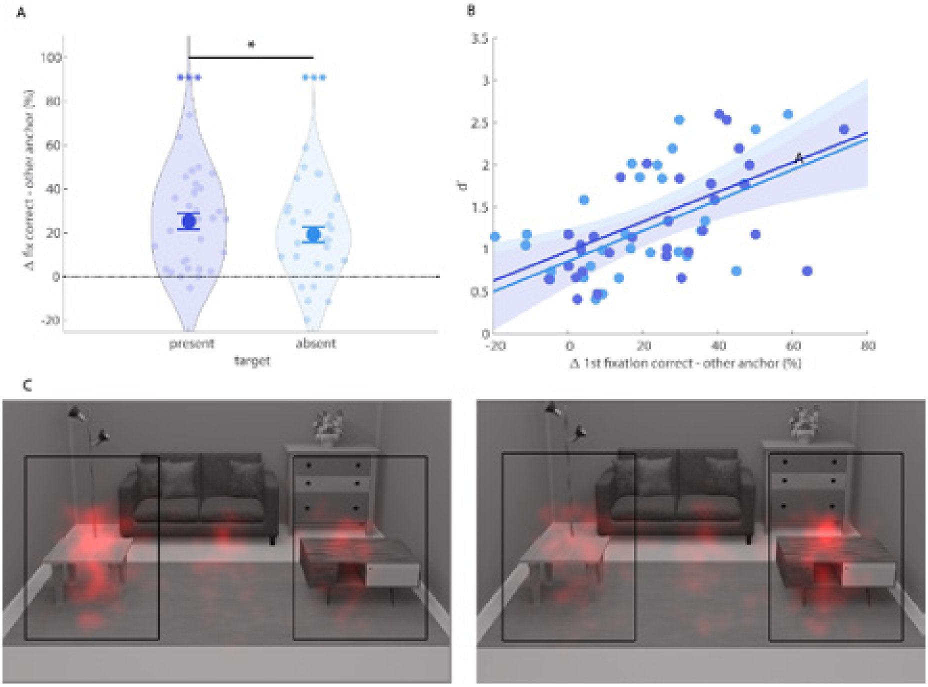

Figure 2 Eye tracking results. A Difference in percentage of first fixations directed towards the correct (relevant) anchor and the other (irrelevant) anchor for target present trials (left) and target absent trials (right). All error bars are SEM. ***, p<.001 **B** Correlations of d' during search with first fixation differences on target present (dark blue) and target absent (light blue) trials across participants. Shaded areas shows 95% confidence intervals. **C** Heatmaps for the first fixations on target absent trials when the associated anchor was on the left (left panel) or right (right panel). The black squares indicate the two areas of interest (AOIs) used for eyetracking analysis and were not visible for participants.

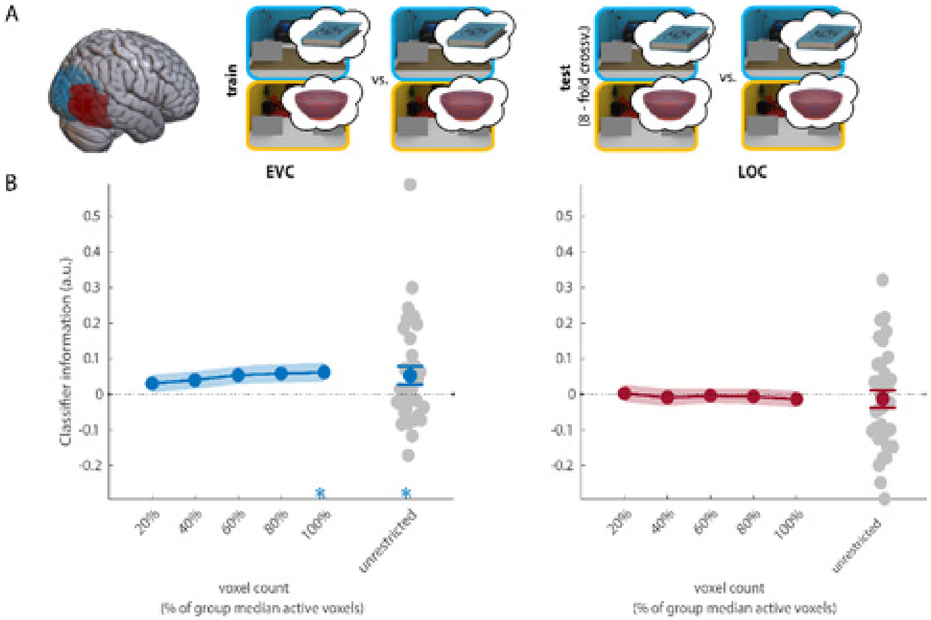

Figure 3: Decoding of the target object within search task runs. **A** Overview of the decoding scheme, decoding the target (book/bowl) from the preview only trials (leave one run out cross-validation). Brain shows anatomical location of the ROIs. **B** Decoding results in EVC and LOC for all selective voxels of individual participants (unrestricted -ROI) and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (unrestricted) ROI. All error bars are SEM. *, p<.05. For sub-ROIs, asterisks show significance after threshold-free cluster enhancement (TFCE).

fMRI results

Preparatory activity in LOC does not reflect the target

Having established reliable guidance by anchors, we turned to investigate preparatory fMRI activity patterns on the preview-only trials within two regions of interest: EVC and LOC. First, we tested whether there was any information about the target object participants prepared to search for (Figure 3A). We trained classifiers to decode the target category (book or bowl) of each trial, collapsed across scene contexts and therefore also across associated anchors. For both ROIs, we intersected group-level masks with voxels sensitive to target and/or anchor features at their retinotopic locations in the task. To ensure the robustness of our results to different voxel selections, we repeated this analyses across ROIs of different sizes (see Figure 3). Classifier performance was calculated based on the distance from the decision boundary, providing a more sensitive and continuous estimate compared to classification accuracy based on binary labels (Walther et al., 2016). This was calculated as follows: $Classifier\ information = \frac{1}{n}\sum_{i=1}^{n} zd_i * l_i$. For each beta *i*, the z-scored distance-to-bound (zd_i) was multiplied by its correct label ($l_i \in [-1, 1]$) and this score averaged across all *n* betas (see Gayet & Peelen, 2022 for a similar approach). All our main results were replicated with accuracy-based analyses (see Figure S3.).

Within EVC, the current search target could be decoded above chance (CI = [0.01, 0.11], p = 0.02; Figure 3B), but this was not consistent across a majority of voxel selections and should therefore be interpreted with caution. Notably, in contrast to previous studies, there was no information about the current search target in LOC (CI = [-0.06, 0.04], p = .58). Taken together, these results indicate that the search target was not strongly represented in preparatory activity during context-guided search. This result is in line with the ineffective attentional guidance offered by target features in this task. As attention (and eye movements) were primarily guided by anchor objects, we expected that preparatory activity would represent these anchor objects.

Preparatory activity for associated anchor objects

Next, we turned to our main analysis, testing whether preparatory activity indeed reflected the target-associated anchor that guided eye movements. For this analysis, classifiers were trained to decode the target-associated table (Table 1 vs Table 2) from fMRI activity patterns in preview-only trials (see Figure 4A). There was no evidence for anchor-information in EVC (CI = [-0.03, 0.14], p = .24; Figure 4B). Importantly, however, the associated anchor could be reliably decoded from LOC (CI = [0.03, 0.15], p = .001; Figure 4B). Decoding was consistent across all ROI sizes. Note that there was no external visual cue for the relevant table and this preparatory activity must therefore have been internally generated, based solely on the learnt target-anchor associations.

Cross-scene decoding analysis: Anchor templates in LOC generalise across contexts

Since the associated anchor could be reliably decoded during search preparation, we next asked whether preparatory anchor activity was independent of the cued target and the specific room (Figure 5A). Thus, we trained classifiers on dissociating the currently relevant anchor in one room, testing whether they could also decode the relevant anchor in the other room, where the target-anchor associations were reversed. This analysis additionally allowed us to rule out that differences in anticipated search difficulty between anchors contributed to anchor-decoding, as the performance differences we observed were scene-specific (see Behavioural analyses in Supplementary Materials).

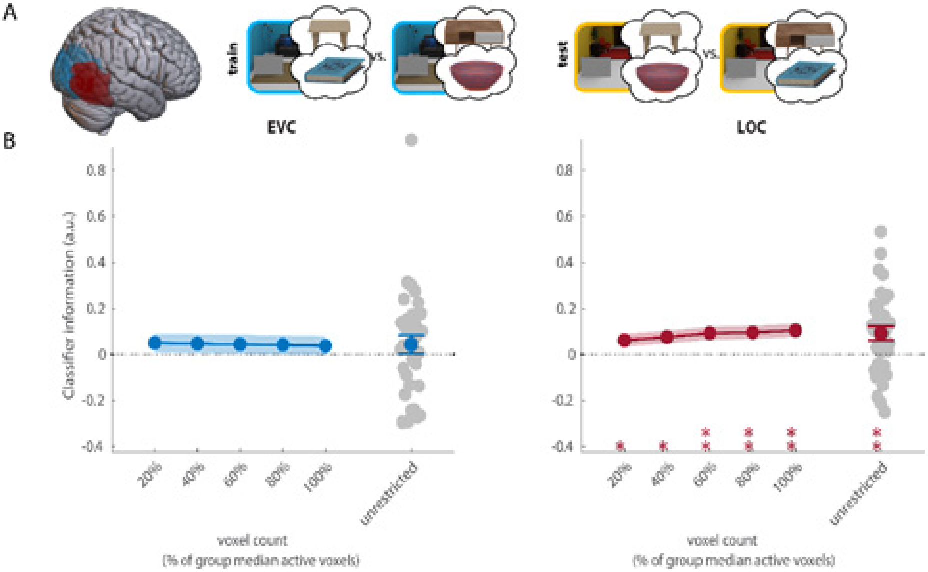

Figure 4: Decoding of the associated anchor within search task runs. A Overview of the decoding scheme, decoding the associated anchor (Table 1/Table 2) from the preview only trials (leave one run out cross-validation). Brain shows anatomical location of the ROIs. **B** Decoding results in EVC and LOC for all selective voxels of individual participants (unrestricted-ROI) and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (unrestricted) ROI. All error bars are SEM. **, p<.01; ***, p<.01. For sub-ROIs, asterisks show significance after threshold-free cluster enhancement (TFCE).

In this analysis, above-chance decoding would indicate preparatory activity for the anchor table, generalising across associated targets, while below chance decoding would indicate preparatory activity for the target (independent of the associated anchor table). Within EVC, preparatory activity did not generalise across rooms (CI = [-0.031, 0.13], p = .28; Figure 5B). Importantly however, the anchor template in LOC was independent of scene context and the associated target (CI = [0.03 0.15], p = .004; Figure 5B). Decoding was consistent across ROI sizes.

Overall, this result shows that preparatory activity in LOC specifically reflected the anchor, irrespective of the associated target and independent of the room in which participants prepared to search.

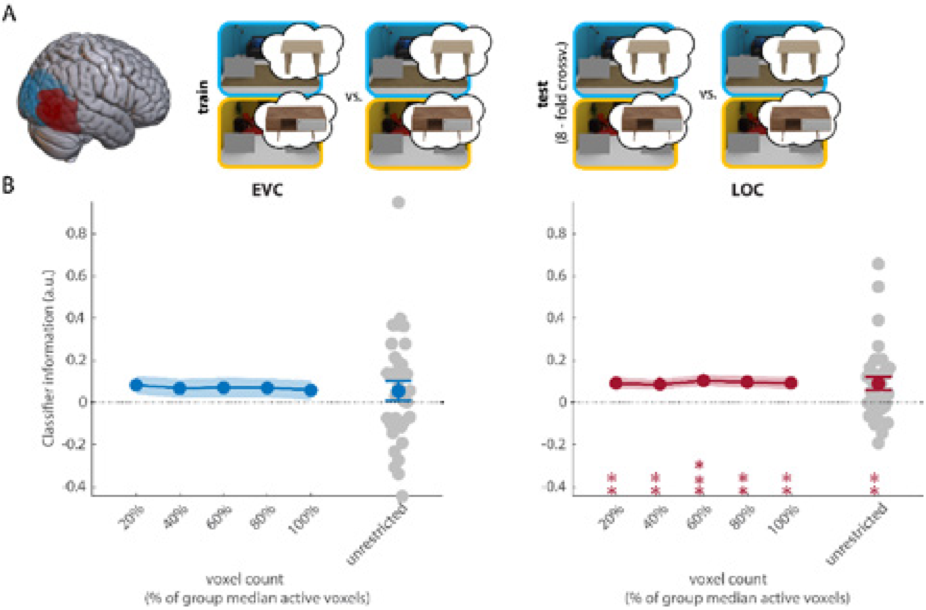

Figure 5: Decoding of the target-associated anchors across rooms. A Overview of the decoding scheme, decoding the associated anchor (Table 1/Table 2) from the preview only trials in one room/ scene context, testing on the other (cross-scene decoding). Brain shows the anatomical location of the ROIs. **B** Decoding results in EVC and LOC for all selective voxels of individual participants (unrestricted-ROI) and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (unrestricted) ROI. All error bars are SEM. *, p<.05; **, p<.01. For sub-ROIs, asterisks show significance after threshold-free cluster enhancement (TFCE).

Searchlight analyses

To complement the ROI analyses, we conducted a searchlight analysis to test whether additional brain regions showed context- and target-independent preparatory activity reflecting the anchor, as we found for LOC. This analysis revealed one cluster with above-chance decoding, located within the right intraparietal sulcus (IPS; Figure 6). We did not observe significant clusters within LOC, likely explained by reduced sensitivity of this searchlight analysis to spatially distributed activity patterns, which may have additionally varied across participants, together with a stringent multiple comparison correction (threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009). Indeed, a whole-brain searchlight analysis at a more lenient threshold (p < .005, uncorrected) revealed clusters of voxels within LOC (Figure S4). There were no significant clusters showing negative cross-scene decoding (which would be indicative of preparatory target activity).

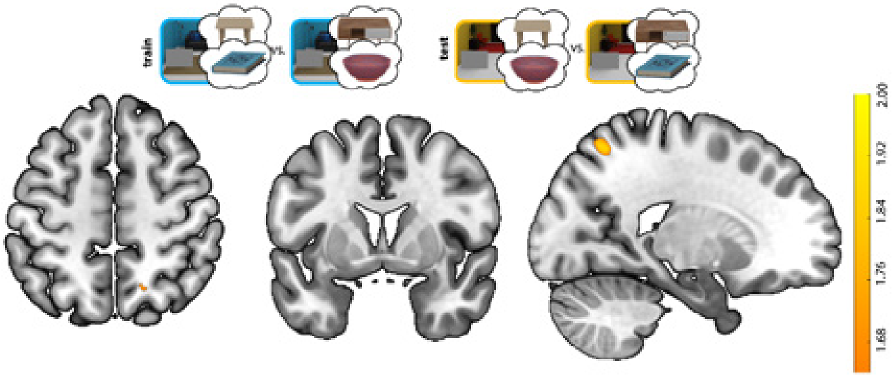

Figure 6: Searchlight for decoding the target-associated anchor across scene-contexts. The searchlight analysis revealed a cluster in the right intraparietal sulcus (x, y, z = 20.1, -63.6, 58, 203 mm^3). Colour indicates z-scores after threshold free cluster enhancement (TFCE).

Discussion

Attention in structured, real-world scenes is often guided by features that are not based on a precise, veridical representation of the target (Wolfe, 2021; X. Yu, Zhou, et al., 2023), and can even be based on entirely different non-target objects, as in the case of guidance by anchor objects. Here, we tested whether and how preparatory activity supports such context-guided search.

Our results show that preparatory activity patterns in LOC, previously implicated in encoding preparatory attentional templates for naturalistic search targets (Gayet & Peelen, 2022; Peelen & Kastner, 2011), reflects relevant guiding objects in the current scene context, rather than the search target per se. Importantly, in contrast to the target object in this or previous studies investigating preparatory activity, the guiding object did not have to be reported, nor was it explicitly cued. Our findings are in line with recent visual search theories (Wolfe, 2021; X. Yu, Zhou, et al., 2023) that propose a dissociation between templates for attentional guidance and templates for later target processing, where a veridical and precise target representation is mostly useful for those later decision stages. By disentangling target and guiding features, our results indicate that preparatory activity in LOC supports guidance itself. Besides target report, the current design also allowed us to dissociate this biasing signal from the external target cue. Such cues are seldomly encountered in real-world searches and the representations based on external cues (even when they are abstract) can differ from internally generated ones, e.g., activate more posterior, purely visual regions (Hu & Yu, 2023). In the current study, the same target cue was associated with different anchors depending on the room, such that the associated anchor had to be retrieved from memory based on learnt associations. Overall, our results show that preparatory activity in visual cortex reflects an internally generated guiding template, independent of external cues or target report.

We found that preparatory activity in visual cortex can be highly flexible and context-dependent, representing a different anchor when looking for the same target in different scene contexts (blue or yellow living room). Across rooms, each target was equally associated with either table; which anchor was effective for guidance depended on the combination of target and scene context. Furthermore, all trials were randomly intermixed in the experiment. The target-anchor associations therefore did not reflect general semantic associations, but rather context-specific associations for guidance, also independent of priming and any visual similarity between targets and anchors. Despite this reversal across contexts, associations were quickly learnt, used for eye movement guidance, and reflected in preparatory activity, thereby demonstrating the flexible top-down nature of preparatory activity. This flexibility would be beneficial for search in daily life, where objects can appear next to different anchors depending on the context (e.g., a coffee cup next to the coffee machine in a kitchen but next to the monitor in an office).

What is the source of those flexible and context-dependent biases in visual cortex? Previous research investigating learnt associations between scene context and target location (e.g., learning that target 1 usually appears in the top left corner of scene 1) has shown that guiding attention based on memory is mediated by interactions between hippocampus and visual cortex (J. J. Summerfield et al., 2006). These interactions have been shown to influence visual cortex activity prior to stimulus onset (Favila & Aly, 2023; Stokes et al., 2012; J. J. Summerfield et al., 2006). Similar mechanisms may be involved when recently learnt context-dependent object associations are retrieved for guidance, as in the current study. It is possible that a different mechanism supports guidance for highly stable object-associations that have been learnt over many years (e.g., toilet roll holder next to toilet) as such associations may affect the cortical representation of the associated objects, leading to a more integrated representation of object pairs within object selective visual cortex (Kaiser & Peelen, 2018).

The searchlight analysis revealed that anchor-specific preparatory activity was also present in the right intraparietal sulcus. Similar to LOC, this activity generalised across target identity and scene contexts. IPS may be well suited to encode such guiding templates, as neurons in this region are shape sensitive (Konen &

Kastner, 2008) and are involved in working memory (Christophel et al., 2015, 2018; Iamshchinina, Christophel, et al., 2021; Rademaker et al., 2019) and eye movement control (Grefkes & Fink, 2005). IPS neurons have also been implicated in "information sampling" (Gottlieb et al., 2013; Gottlieb & Oudeyer, 2018), encoding the expected information gain of attending to a particular object (Foley et al., 2017; Vossel et al., 2006, 2015). Within naturalistic scenes, both IPS and LOC have been found to encode expected spatial locations of targets, thereby supporting contextual guidance in scenes (Preston et al., 2013). This could suggest that these regions are sensitive to both the location and the identity of guiding objects. Future research targeting IPS and its subregions more specifically is needed to confirm these findings.

While we found no evidence for preparatory activity related to the target object in LOC, participants must have correctly prepared for the cued target, as they performed above chance and target features improved guidance. It is possible that such activity starts later, once the associated anchor has been located and participants need to decide whether the search target was present or absent. In the current experiment, this stage could not be separately investigated. In line with this, there is evidence from sequential search and working memory tasks that items relevant in the future are not (or less well) decodable in visual cortex, compared to immediately relevant items (e.g., Christophel et al., 2018; van Loon et al., 2018). Future studies could use our paradigm in combination with time-sensitive methods (e.g., EEG/MEG) to test for sequential anchor-target activity. Alternatively, the current paradigm could be adjusted for fMRI to measure preparatory activity after the anchor has been located, for example using a two-step preview task. More generally, while other, more sensitive designs or analyses might reveal concurrent target and anchor templates, this would not change our main conclusion that preparatory activity can represent non-target features based on learnt associations.

Our results support the view that attentional templates, especially guiding templates, are flexible and represent useful guiding information rather than necessarily a veridical and detailed representation of the target (X. Yu, Zhou, et al., 2023). Whether a guiding template reflects target or non-target features is most likely a trade-off, depending on the relative ease with which target features themselves can be effectively used for guidance (influenced by, e.g., saliency, predictability, target-distractor similarity), and the degree to which non-target features are associated with the target (making them more or less informative). In this study, our aim was to test whether preparatory activity reflected such non-target features in situations where these provided an advantage for guidance. If guidance by target features was relatively more effective, e.g., because target features were more salient or target-anchor associations were unreliable, it is likely that preparatory activity would more strongly reflect the target, as reported previously (Chelazzi et al., 1993; Gayet & Peelen, 2022; Peelen & Kastner, 2011; Stokes, Thompson, Nobre, et al., 2009). Testing this trade-off in detail remains a relevant question for future research, as object associations in the real world vary in strength across target-anchor pairs and are not always deterministic. At the same time, it is likely that anchor templates in real-world search are relatively reliably activated during search because of our extensive experience with real-world object associations (Kallmayer et al., 2024).

The current study allowed us to isolate one key aspect of naturalistic search - guidance by target-associated objects in the scene - and test the content of preparatory activity using a controlled design. The tables in our task were similar to real-world anchors in terms of being large and salient objects that provided spatial predictions about target locations, thereby constraining search. However, our study does not capture all aspects of real-world anchor search. For example, real-world target-anchor associations are typically learnt over many years (e.g., toothbrush-sink) and are characterised by high semantic, functional, and/or contextual similarity of the objects involved, which are difficult to tease apart in an experiment. Besides guidance by anchor objects, naturalistic scenes also contain a wealth of other regularities that can be quickly extracted and support search (Hollingworth, 2012; Peelen & Kastner, 2014; Võ, 2021; Wolfe, Alvarez, et al., 2011). For instance, scenes provide contextual expectations about target appearance that can shape the preparatory template in visual cortex (Gayet & Peelen, 2022). Target location may also be predicted based on the overall spatial layout of the scene, guiding attention and eye movements (Castelhano & Heaven, 2011; Neider & Zelinsky, 2006). Furthermore, frequently co-occurring objects may be grouped together, reducing competition between objects in the scene when those objects act as distractors (Kaiser et al., 2014). Rather than fully recreating all aspects of real-world search, the current findings reflect a proof-of principle of a key aspect of context-guided search, showing how preparatory activity can support context guidance based on learnt object-associations.

In conclusion, we show that context-guided search is supported by self-generated and context-dependent preparatory biases in object-selective cortex, clarifying the functional role of preparatory activity during visual search. The ability to quickly learn novel object associations and flexibly adapt preparatory biases to represent the most useful features for attentional guidance is an important factor contributing to the efficiency with which we select and interact with relevant objects in our daily life environments.

Materials and methods

Participants

Thirty-four participants (22 women, mean age: 23.5, sd: 5.8) from the Radboud University subject pool were included in this study. Seven additional participants were tested, but not included in the final data set: five due to at-chance accuracy in the search task (determined by a one-sided binomial test with $\alpha = .05$), one chose to abort the experiment and one due to excessive head movement. Those participants were replaced until reaching the planned sample size of 34, sufficient to uncover an effect of medium size ($d = 0.5$) with 80% power. All participants received 25€ for their participation and provided written informed consent, declared themselves free of epilepsy, and had normal or corrected-to-normal vision. All procedures were approved by the local ethics committee (CMO region Arnhem-Nijmegen, the Netherlands, Protocol CMO2014/288).

Search task

Participants completed 256 trials of the task in the fMRI scanner (eight runs of 32 trials each). They searched for two different types of targets (books or bowls) in two 3D rendered living room scenes (see Figure 1).

Each trial began with a target cue (‘Book’ or ‘Bowl’) written on screen (1s), followed by a living room preview, with grey occluder rectangles placed at the location of the tables, slightly below fixation and in the left and right periphery. During target cue and scene preview, participants were asked to keep their gaze on a central fixation cross. The preview remained on for 4.4s in preview-only trial (128 trials) or was replaced by a search scene (400 ms, remaining 128 trials). For the search trials, a response screen then prompted participants to indicate whether the target was present or absent by using a button box with their right hand (max response duration 1.5 s). Feedback was provided after every search trial by a red or green dot at fixation. After every run, participants were additionally given feedback on their average performance.

Target category and scene context changed on a trial-by-trial basis and the factors Trial type (search, preview only), Target presence, Target category (Book, Bowl), Scene context (blue, yellow living room) and Position of the associated anchor (left/right) were all counterbalanced within each run.

Localiser task

To localise voxels that were selective to the targets and/or anchors, we also included four runs of a separate localiser task. In this task, participants saw the anchor and target objects used in the main task, now presented in isolation on a grey background. There were either two identical objects shown in the periphery (at the location of the table centres in the search task) or a single object in the centre of fixation (as they would appear after an eye movement to fixate them). Each run consisted of 32 miniblocks (four repetitions of eight conditions: book/bowl/Table 1/ Table 2 x central/ peripheral location).

Within each miniblock, participants saw eight stimuli of the same category (0.4s on, 0.6 s off, 2s fixation baseline after each miniblock) and had to respond to size-oddball targets, in which the stimuli appeared slightly larger than usual (i.e., deviating from their typical, memorised size; 1.15x size increase for anchor tables, 1.5x for target objects) by pressing a button. There was one such oddball target per miniblock and participants received feedback about the percentage of found targets after each block. For target miniblocks, different exemplars were shown within a miniblock. As this was not possible for the anchor tables, we introduced slight rotations along their vertical axes in the images, to avoid a pixel-based strategy for the task and decrease repetition suppression effects.

The localiser task was presented interleaved with the search task, with one localiser followed by two search task runs. For the first three subjects, a different localiser task was used. While they saw the same stimuli (without oddball size targets), they were asked to respond to a brief rotation of the fixation cross. All other presentation parameters were the same for all participants.

Of note, for all analyses, we did not observe generalisation (i.e., successful cross-decoding) between visually evoked responses in this localiser and the preview trials of the main task. This may be due to insufficient statistical power to detect such a generalization or that different features were relevant to discriminate objects in the periphery for quick guidance of eye movements and comparing objects to their memorised size in this localiser task, resulting in less overlap across tasks (Henderson et al., 2023; McKee et al., 2014). Another possibility is that preparatory activity in our task was overall less sensory-like (Gong et al., 2022). As such, this null-finding is difficult to interpret, and our analyses focus on cross-validation across runs of the main experiment and cross-decoding across different scenes, agnostic to the representational format of the preparatory template.

Anchor-target association training

Immediately prior to scanning, participants practiced the search task and learnt to associate anchors and targets. They saw example target objects and search scenes and practiced the search task (32 trials). During this practice they were informed that within each of the two rooms, each target was associated with one of the two tables in the room, and that it would always appear on the respective table if it was present.

After this practice, participants were tested on the associations using a 2AFC-task (12 trials). On each trial, they were shown one of the living rooms with empty tables left and right and the name of a target written in the image centre, and had to indicate the location of the associated table given target and scene-context using a keyboard. Performance on this task was 83.09% (sd 14.72). Participants received feedback after each trial and were reminded of the associations after the task if needed.

Stimuli

Six hundred forty 3D rendered search scenes were created using Blender 2.92 (Blender Foundation) and the Eevee rendering engine. These scenes depicted a blue or yellow living room with identical spatial dimensions and camera positions, but different furniture.

In both rooms, two visually distinct tables were placed in the foreground, slightly below fixation, on the left and right. Four objects were placed on each table, each being either one of the target objects (books and bowls, drawn from 20 unique exemplars per target category, varying in colour and size), or non-target objects that could be plausibly found on living room tables (40 objects, e.g., a tea mug, chess board, baseball cap, headphones, game controller, vase, candles, laptop). Three hundred twenty unique object constellations were created - 20 exemplars for each of combination of counterbalancing factors (scene context, target presence, associated anchor left or right) and the associated anchors (counterbalanced across participants) added. Models for furniture, targets, and distractor objects were taken from https://sketchfab.com/ or newly built/modified. Custom python scripts ensured objects did not intersect with each other, and images were manually checked to ensure objects did not significantly occlude each other. Search scenes subtended about 25 x 12.62° (20 x 11.25° for the first three participants).

Scene images with blurred tables (both tables averaged and pixelated left and right) were used as masks after the search scene, and two scenes with grey occluder squares in front of the tables used as preview.

For the localiser, isolated target and anchor objects were rendered without the scene background, either at the peripheral locations the anchors occupied in the search scene or in the centre at the same height. For the target objects, 20 images per localiser condition showed all category exemplars used in the search task. For the anchor objects, ten images were created per condition, showing the same table slightly differing in its rotation along the vertical axis (equidistant steps from -10 to 10 degrees from their typical orientation in the search scenes).

Setup & eye tracking

Inside the scanner, stimuli were presented on a 32 inch IPS BOLD screen (1920 x 1080 pixels, 120 hz refreshrate) placed at the back of the scanner bore, and could be viewed by participants through a mirror mounted on the head coil.

During scanning, participant’s left eye was tracked by an Eyelink 1000+ eye tracker (SR Research, sampling rate 1000 Hz). At the start of the experiment, the eyetracker was calibrated using a nine-point calibration procedure (resorting to a five-point calibration in case no accurate nine-point calibration was possible) and recalibrated if required.

Eye tracking AOIs

Three rectangular areas of interest (AOIs) were defined for analysing fixations during the search scene. First, a centre AOI around the fixation cross (4.5 x 10.5°) to ensure fixation was in the image centre before the search scene appeared and one centred on the left and right anchor table each (7.5 x 10.5°). On 94.15% (sd 9.65) of trials, at least one fixation was recorded during presentation of the search scene, and 27.08% (sd 10.32) of trials had two or more recorded fixations. 83.33% (sd 19.36) of trials with a fixation also had an initial fixation in the centre AOI at scene onset, of which 80.52% (sd 10.76) of first fixations were then directed at one of the anchor AOIs.

fMRI acquisition and preprocessing

Data were acquired on a 3T Siemens SKYRA Scanner using a 32-channel head coil. A T2-weighted gradient echo EPI sequence was used for acquisition of functional data (TR 1.5 s, TE 33.4 ms, flip angle 75°, 2 mm isotropic voxels, 68 slices, 4x multiband acceleration factor). For the search task, 198 images were acquired per run and 235 images were acquired per run for the localiser runs. A high-resolution

T1-weighted anatomical scan was acquired prior to the experimental runs, using an MPRAGE sequence (TR 2.3 s, TE 3.03 ms, flip angle: 8°, 1 mm isotropic voxels, 192 sagittal slices, FOV 256 mm).

Data preprocessing was performed using SPM12 (https://www.fil.ion.ucl.ac.uk/spm/). Preprocessing steps included spatial realignment, co-registration of functional and anatomical scans and normalization to MNI 152 space. A Gaussian filter (FWHM 3 mm) was then applied to smooth the images.

Subject level GLMs were estimated on the pre-processed images. For the search task, the four combinations of targets and scene context were modelled as regressors of interest, by convolving a boxcar function spanning from preview onset to offset with the canonical HRF curve provided in SPM12. Importantly, this model only included preview-only trials.

For the localiser runs, individual miniblocks were modelled as boxcar functions spanning the duration of a miniblock and were also convolved with the canonical HRF curve. There were 8 conditions of interest: books, bowl, anchor Table 1 and anchor Table 2 each presented centrally or peripherally.

Six motion regressors and one run-based regressor were included as nuisance regressors and all betas were estimated on a run-based basis.

ROI definition

Two main ROIs (EVC, LOC) were defined for each participant. ROI definition was based on group level masks (EVC: Brodmann areas 17, 18, LOC: taken from Julian et al. (2012)), resliced into MNI space. To exclude voxels of non-interest (e.g., white matter voxels, visually unresponsive voxels), these group-level masks were intersected with contrasts from the localiser task for each participant, as follows: within each ROI, we selected voxels that discriminated between the targets or anchor objects, when they were visually presented during the independent localiser task, by intersecting the group levels masks with respective F-contrasts (book vs. bowl or Table 1 vs. Table 2). This was done separately for each hemisphere, to be sensitive to potential hemispheric differences. For each contrast, hemisphere and brain area, we first created one large ROI mask (unrestricted -ROI) including all active voxels based on that contrast ($p < .05$, uncorrected) by intersecting the respective group-level masks and participant contrast maps. In a last step, we combined the masks created for each contrast, thus including both anchor and target-discriminating voxels within one ROI-mask, and results are based on these combined ROI maps. Averaged across participants and hemispheres, these masks included 1289.85 voxels in EVC (sd 58.17) and 917.32 voxels (sd 50.72) in LOC. This procedure ensured we included informative voxels, in which any preparatory bias for targets or anchors should be strongest, while still basing our analyses on a large enough subset of voxels.

To ensure that our results were robust to different ROI definitions, we also created smaller sub-ROIs for each brain area and participant. Five sub-ROIs were created, including an equal number of the top x most target and anchor selective voxels within each hemisphere in equidistant steps up to 424 voxels for EVC and 264 voxels per hemisphere for LOC. Those numbers reflected the median number of significantly target-selective voxels across hemispheres and participants (always lower than the number of anchor-selective voxels). All our main results were also consistent when considering the target- or anchor-selective voxel ROIs in isolation (see Figure S1).

Multivariate pattern analysis

All multivariate analyses were performed using linear support vector machines (SVMs), using The Decoding Toolbox (TDT: Hebart et al., 2015) and the libsvm library. Classifiers were trained and tested on the run-based beta-weights after GLM estimation. For every ROI and voxel count, separate classifiers were trained for each hemisphere. As we did not observe any hemispheric differences, the final classification performance was then averaged across hemispheres before statistical testing (see Figure S2 for decoding results within individual hemispheres).

Classification was either performed in (1) a leave-one-run out eight-fold cross-validation scheme to test for target or target-associated anchor information within the search task runs (training the classifier to, e.g., distinguish all book and bowl trials), or (2) a cross-classification scheme to test for generalization across scene contexts (training on search task preview trials of one scene context and testing on the other). For cross-classification, the classification directions were averaged before statistical testing.

Searchlight analysis

To complement the previous ROI analyses, we ran an additional searchlight analysis, testing for generalization of the anchor template across scenes (one-sided test). Similar to the ROI analyses, we first identified target- and/or anchor-selective voxels by running additional searchlights (restricted to a mask of cerebral cortex regions based on the AAL atlas (Tzourio-Mazoyer et al., 2002) to decode the targets or anchors from the localiser betas. Voxels with significant decoding for either the target or anchor at the group-level ($\alpha = .05$) were then used to create a mask for the cross-decoding searchlight. For all searchlights, spheres had a radius of five voxels (resulting in around 524 voxels in total).

Statistical tests

To test for differences between conditions in the behavioural/eye tracking analysis or to determine above-chance decoding in the all-ROIs, we performed bootstrap tests against chance, resampling individual participants with replacement for 10000 iterations. P-values reflect two-tailed tests unless indicated otherwise.

Threshold free cluster enhancement (TFCE) was applied to the sub-ROI data and searchlight analyses, using the CoSMo MVPA toolbox (Oosterhof et al., 2016). This procedure boosts belief in neighbouring data points containing signal, and was applied both to real as well as synthetic null-data. The final p or z-values reflect how likely a given TFCE score is, given the maximum TFCE scores across the null-data and thereby accounts for multiple comparisons.

Supplementary materials

Behavioural analyses

Overall behavioural accuracy in the search task was 72.46% (sd 8.91). Hit rate was 66.16% (sd 12.61), with 21.16% (sd 10.89) false alarms. Sensitivity (*d′*) was 1.51 (sd 0.61) and criterion (*c*) 0.21 (sd 0.24) on average. Sensitivity (*d′*) was 1.51 (sd 0.61) and criterion (*c*) 0.21 (sd 0.24) on average. Thus, as intended, finding the targets in the search task was challenging, but not impossible.

We established whether performance was matched for both targets and scene contexts, using a Target (Book, Bowl) x Scene Context (yellow, blue living room) ANOVA. Participants were equally accurate for both targets and both rooms (no main effect of Target: $F(1,33) = 1.04$, $p = .32$, $\eta_p^2 = 0.03$; no main effect of Scene context: $F(1,33) = 0.07$, $p = .79$, $\eta_p^2 < .01$; no Target x Scene context interaction: $F(1,33) = 3.06$, $p = .07$, $\eta_p^2 = .09$). Next, we tested whether performance differed for search on the two anchor tables, using an Anchor (Table 1, Table 2) x Scene Context (yellow, blue living room) ANOVA. Participants were more accurate on the smaller and brighter table ($F(1,33) = 12.85$, $p < 0.01$, $\eta_p^2 = 0.28$), possibly due to the better contrast between the table and the objects placed on it. However, this advantage depended on the context, as indicated by a significant Anchor x Room interaction ($F(1,33) = 5.84$, $p = .02$, $\eta_p^2 = .15$). While performance for the two tables differed in the yellow living room (CI = [4.86, 10.85] $p < .001$) they did not differ for the blue living room (CI = [-3.2, 5.19], $p = .58$). Altogether, these findings indicate that there were no consistent accuracy differences across targets, anchors, or rooms.

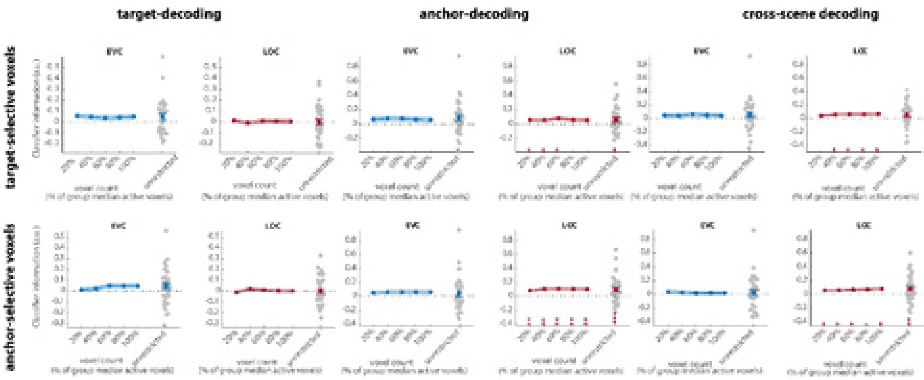

Figure S1: Decoding results shown separately for target and anchor-selective voxels. Decoding in EVC and LOC for all selective voxels of individual participants (unrestricted-ROI) and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (unrestricted) ROI. All error bars are SEM. *, p<.05; **, p<.01

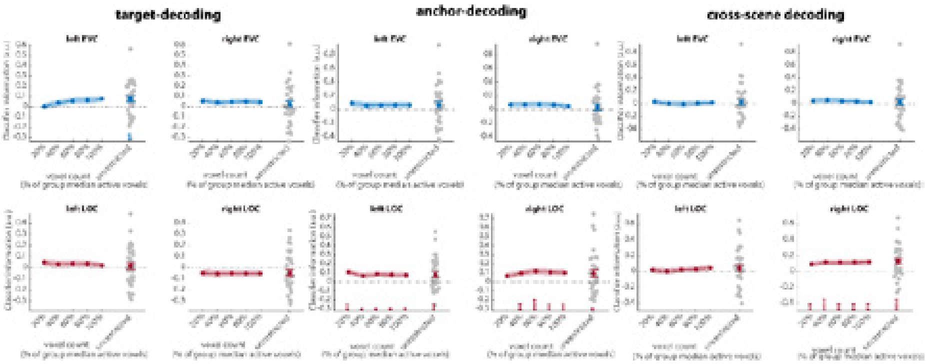

Figure S2: Decoding results shown separately for left and right hemisphere. Decoding in EVC and LOC, for all selective voxels of individual participants (unrestricted-ROI), and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (all) ROI. All error bars are SEM. *, p<.05; **, p<.01;***, p<.001

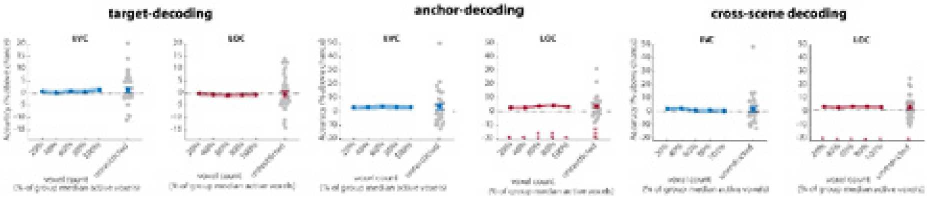

Figure S3: Decoding results using decoding accuracy. Decoding in EVC and LOC for all selective voxels of individual participants (unrestricted-ROI), and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (all) ROI. All error bars are SEM. *, p<.05; **, p<.01; **, p<.001

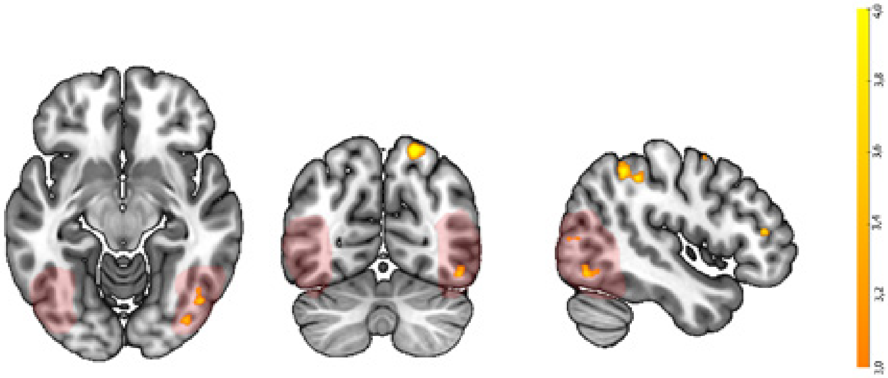

Figure S4: Uncorrected whole-brain searchlight (no voxel selection) for decoding the target-associated anchor across scene-contexts. The searchlight shows voxel clusters within the LOC mask (marked in red). Searchlight is thresholded at p = .005 (two-sided). Colour indicates t-values without threshold-free cluster enhancement (TFCE).

Chapter 4

Blinded by the Mind's Eye? Investigating Competitive Interactions between Imagery and Perception

Visual mental imagery relies on the top-down activation of visual cortex representations, representations that are also activated during the perception of external visual stimuli. These internally and externally generated signals are often simultaneously present in visual cortex. Given that both imagery and perception draw on similar and capacity-limited neural resources, we hypothesised that they compete for representation, mutually degrading their individual representations as a function of their representational similarity. To test this hypothesis, we decoded multivariate EEG activity patterns evoked by visual stimuli while participants (N=34) engaged in vivid visual imagery of naturalistic scenes containing objects from one of two domains (buildings or animals), each drawn from one of four categories. During the imagery delay, task-irrelevant intervening objects appeared, belonging to either the same domain (and potentially the same category) or the other domain. We densely sampled (eight objects per trial) neural responses to intervening visual stimuli, comparing the responses to visually identical object stimuli as a function of matching or mismatching imagery content. This allowed us to separately assess imagery-perception competition at the category and the domain-level. Results showed that the representations of the identity and the location of intervening stimuli were degraded within 200 ms after stimulus onset when participants were imagining an object of the same category. Imagery-related interference was specific to the category-level. These results indicate that imagery interferes and competes with the perceptual representation of categorically similar objects, shedding light on interactions between internally and externally generated representations in visual cortex.

Introduction

While navigating and perceiving our surroundings, we are often not only processing external stimuli, but also attend to our internal representations at the same time. For instance, I may be walking around in my room listening to an audiobook and vividly imagining the characters and places described, while still having a sense of my surroundings (e.g., seeing my cat entering the room).

For most people, seeing images with their mind's eye resembles perception, an intuition supported by neuroimaging evidence. Imagery activates category-selective areas in the ventral visual cortex (Ishai et al., 2000; O'Craven & Kanwisher, 2000). Furthermore, highly similar multivariate activity patterns are evoked by seeing and imagining a given object (Albers et al., 2013; Cichy et al., 2012; Dijkstra et al., 2017; Naselaris et al., 2015; Stokes, Thompson, Cusack, et al., 2009). Recently, the overlap between imagery and perception in the visual cortex has been observed even at the single-neuron level (Wadia et al., 2024), indicating that imagery relies on top-down activation of visual representations (Dijkstra et al., 2019; Pearson, 2019). Given this overlap, it is important to establish how representations of perceived and imagined stimuli interact and co-exist in visual cortex.

Previous studies have shown that imagery can facilitate processing of upcoming external stimuli, where holding an image of a task-relevant object in mind can provide a competitive advantage for matching stimuli (Cochrane et al., 2021; Dijkstra et al., 2021; Farah, 1985; Ishai & Sagi, 1995; Moriya, 2018; Pashler & Shiu, 1999; Pearson et al., 2008; Sulfaro et al., 2024), facilitating their detection and providing attentional guidance towards their location. Importantly, however, co-existing representations of imagined and perceived objects may also compete when both are drawing on similar and limited neural resources (Sulfaro et al., 2023). Broadly in line with this, concurrent visual stimulation can impair imagery strength (Keogh & Pearson, 2011, 2014), and, conversely, imagery can impair visual detection performance (Craver-Lemley & Arterberry, 2003; Craver-Lemley & Reeves, 1992; Finke, 1986; Ishai & Sagi, 1987; Schorr et al., 1978; Segal & Fusella, 1970). At a neural level, internally generated representations in early visual cortex deteriorate, or can even be entirely wiped out, by ongoing visual stimulation (Bettencourt & Xu, 2015; Lorenc et al., 2018; Rademaker et al., 2019; Rademaker & Serences, 2024).

Competitive interactions have been well studied for multiple external stimuli (Desimone, 1995; Reynolds & Heeger, 2009), also for higher-level object representations that show the strongest overlap between imagery and perception (Breedlove et al., 2020; S. H. Lee et al., 2012). When several objects are simultaneously presented and processed, they compete for neural representation. Their individual representations in object- and category-selective regions of the ventral stream are degraded (e.g., reflected in worse multivariate decoding for these objects) compared to when they are presented in isolation (MacEvoy & Epstein, 2009; Reddy & Kanwisher, 2007; Zoccolan et al., 2005). The degree of mutual interference between external objects is a function of their representational similarity. Objects close-by in representational space will co-activate similar neural populations, thereby reducing the quality of individual object representations. On the other hand, objects with highly distinct representations, drawing on distinct neural subpopulations, do not strongly interfere with each other (Bao & Tsao, 2018; Cohen et al., 2014; Franconeri et al., 2013; Kliger & Yovel, 2020; Yeh et al., 2024). If imagery indeed activates the same object representations within visual cortex as perception, similar content-specific competition may be expected when one of the two objects is imagined. However, for competition between internal and external stimuli, the relationship between representational similarity and competition may not be fully linear. One study found that internal representations (held in visual working memory) can facilitate perceptual processing of directly matching external stimuli but compete with mismatching objects that are nearby in feature-space (Kiyonaga & Egner, 2016). Overall, while it is generally established that ongoing external stimulation can have a detrimental effect on internal representations (Bettencourt & Xu, 2015; Lorenc et al., 2021; Rademaker et al., 2019; Rademaker & Serences, 2024), it is an open question whether internal representations can interfere with perception depending on the similarity of their representations, and especially neural evidence for such similarity-based competition is lacking.

Here, we directly investigated whether and when during perceptual processing concurrent imagery impairs the representation of visually presented objects using EEG. Specifically, we asked participants to imagine visual scenes containing one of eight different real-world categories (mammals, birds, fish, reptiles, churches, bridges, bungalows and lighthouses), taken from the domains of animals and buildings, as vividly as possible. While participants were engaged in imagery, a series of intervening objects appeared. Using multivariate pattern analyses, we investigated the representations of the intervening stimuli at different processing stages, decoding their location, as well as their category. To evoke strong competition, the imagined and perceived objects were selected to evoke similar but not fully overlapping object representations in higher-level visual regions (e.g., using different exemplars and subcategories, as well as different locations for the intervening objects). In addition, intervening objects were task-irrelevant to ensure competition arose at the representational level, rather than at the level of task or response.

To test for competitive interactions, we compared decoding accuracy for visually identical intervening objects, while participants concurrently imagined either a similar object (matching the general domain and potentially category) or dissimilar object (mismatching the general domain). We hypothesised that imagery-perception competition would result in a worse representation of similar than dissimilar objects. Our design allowed us to assess competition at different levels of similarity. First, we tested for competition at the general domain level (buildings vs. animals). Categories within the same domain lie closer together in representational space than categories across domains (Konkle & Caramazza, 2013; Kriegeskorte, 2008; Sha et al., 2015). Thus, imagining a building should interfere more with the perceptual processing of another building compared to an animal (and vice versa for imagining animals). Second, we tested for competition at the finer-grained category level. While objects of the same domain would be represented more similarly to each other than to categories of another domain, there can still be a considerable amount of heterogeneity within a domain. Especially within the animate domain, representations of different animal categories, such as mammals and reptiles can still be relatively distinct (Connolly et al., 2012; Sha et al., 2015; Thorat et al., 2019). Thus, while we may observe competition within a domain (e.g., between mammals and reptiles), we hypothesised that competition at the category level should be strongest, as objects of same category lie closest in representational space. To test this, we analysed imagery competition separately for the imagined category and other categories taken from the same domain, thereby distinguishing between category- and domain-level competition.

In line with our hypotheses, we found that imagery interfered with the perceptual processing of intervening task-irrelevant stimuli, specifically when intervening stimuli matched the category of the concurrently imagined stimulus.

Methods

Participants

Thirty-four participants (five male) were recruited from the Radboud University participant pool (SONA Systems) and took part in the experiment. This number was chosen a priori, as it allowed to detect a medium sized effect (*d* = 0.5) with 80% power. One additional participant was tested but had to be replaced due to a technical issue with the EEG system and was not included. Individual age was not recorded, but all participants were between 18-35 years old based on prescreen restrictions. All participants provided informed consent, declared themselves free from epilepsy and all procedures were approved by the Radboud University Social Science Ethics Committee (ECSW-2022-079). Participants received either course credit or 20-30€ for their participation.

All participants were able to create visual mental images, as indicated by their scores on the Vividness of Visual Imagery Questionnaire (VVIQ Marks, 1973). This questionnaire consists of 16 items, asking respondents to visualise different images such as a familiar friend, the front of a frequently visited shop or a sunset at the beach, rating their vividness for each item from 1 (no image at all) to 5 (perfectly clear and vivid as normal vision) and scores thus range from 16-80. The mean VVIQ score across participants was 63.74 (range 41- 80).

Imagery task

On each trial of the retrocued imagery task (see Figure 1A), participants were first presented with two zoscene photographs, one containing an animal, the other a building. Both appeared sequentially within a white square in the centre of the screen (1s presentation each, with 0.25 s in between). Stimuli were taken from eight separate categories: mammals, birds, fish, reptiles, lighthouses, bungalows, churches and bridges (see Figure 1B). Following these two images, a number retrocue indicated whether the first or second image had to be imagined as vividly as possible during the subsequent delay period with open eyes.

During this delay period, a series of eight grayscale images, taken from the same eight categories, briefly appeared around the outer corners of the central square (see Figure 1C). Each of the intervening images was shown for 150 ms, with randomly jittered onset but at least 300 ms between subsequent images (mean 368 ms , maximum 850 ms). These intervening stimuli appeared on 80% of trials. Intervening stimuli were task-irrelevant, participants were told to focus on their imagery while these objects appeared. A given intervening image could match or mismatch the imagined domain (seeing a building when imagining either a scene with a building or an animal), and also at the category level (matching either the cued or uncued category of the trial). They were, however, always different exemplars from the objects shown within the imagined scenes, often taken from different subordinate levels, i.e., imagined mammals were a wolf, a camel, a bear and a boar while the intervening mammal stimuli were a puma, a sheep, a horse and a cow. This ensured that the imagined images evoked similar (and competing) representations within high-level visual cortex, but reduced overlap of low-level features. Each category was shown once during the imagery delay (in random order), equally distributed across the four quadrants.

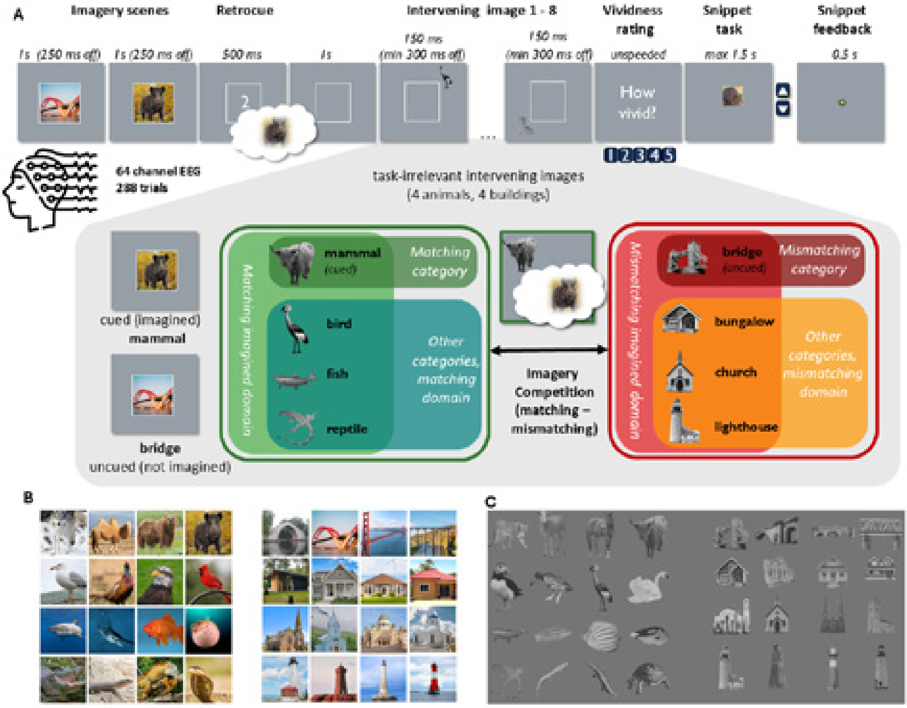

Figure 1: Task Design and Stimuli. **A** Illustration of the imagery task. EEG was measured during a retrocued imagery task. On each trial, participants first saw two scenes (showing either an animal or a building). The retrocue indicated whether the animal or building had to be imagined in the subsequent imagery delay. During the delay, eight intervening images were presented while participants engaged in imagery. These intervening images could match or mismatch the imagined scene at the domain and/or category level. Following imagery trials, participants were prompted to rate the vividness of their imagery on the current trial and perform a snippet task on 50% of trials, in which they had to discriminate between a small cutout of the imagined scene or from a close foil image. The representations of the intervening images were analysed as a function of their overlap with imagery. The intervening image could be taken from the same domain as the imagined scene (matching domain) or the other domain (mismatching domain). In addition to domain-level match, the intervening object could also match either the cued or uncued image of the trial at the category level (matching or mismatching category), or be taken from another category within either the same domain as the imagined image (other categories, matching domain) or the other domain (other categories, mismatching domain). **B** Overview of all imagery scenes, representing mammals, birds, fish, reptiles, bridges, bungalows, churches and lighthouses. **C** Overview of all intervening images, taken from the same categories as the imagery scenes.

Following the imagery delay, an untimed prompt appeared, instructing participants to rate the vividness of the imagery on the current trial, using the number keys (1-5) on the keyboard. Afterwards, participants completed an additional snippet task on 50% of trials, included to ensure participants conjured a vivid and detailed visual image of the imagery scene. A cutout of either the previously imagined scene or a very close foil image (showing a highly similar animal/building or the same building from a slightly different viewpoint) was presented and participants were asked to indicate whether this image was the imagined scene or the foil by pressing the "up" or "down" arrow. A green, red or yellow fixation dot appeared afterwards, indicating a correct, incorrect or late (over 1.5s) response to the snippet task before the next trial started. At the end of each experimental block, participants also received feedback on their average accuracy in this snippet task.

On an additional 1/3 of trials, the retrocue showed a zero, indicating that nothing had to be imagined. Intervening objects were still presented on these trials, but this was not followed by vividness ratings or a snippet task. These trials were not further analysed. Our main question focused on content-specific competition by imagery, and the no-imagery trials were not fully matched to the matching and mismatching imagery conditions in terms of task demands, attentional focus and eye movements and are therefore not further reported.

There were 288 trials in total (eight blocks with 36 trials each). Imagery cue (animate, inanimate, no cue (not analysed)) and the order of imagery scenes (imagining first or second scene) were fully counterbalanced within each block, imagined category and exemplars (four per category) were counterbalanced across the entire experiment. The cutout for the snippet task (four cutouts of the imagined and foil scene each) and the exemplars, order and location of the eight intervening stimuli (two stimuli appearing in each of the four quadrants around the central box) were randomised but equalised as much as much possible across the entire experiment.

Stimuli

The imagery stimuli consisted of 32 coloured images showing mammals, birds, reptiles, fish, lighthouses, churches, bungalows and bridges embedded in their natural scene context (see Figure 1B). Each of these imagery stimuli was also paired with a very close foil-image. All animal photographs and their foils were taken from the THINGS database (Hebart et al., 2019), and building photographs taken from different online sources.

For the intervening stimuli 32 images showing isolated exemplars from the same categories were taken from various online sources, some animal stimuli overlapping with those used in Shang et al. (2024), and converted to grayscale (see Figure 1C).

Imagery scenes were shown within a 7.4° square, intervening stimuli subtended 3.5 x 3.5° and appeared around the four corners of the central square.

Procedure and Setup

Scalp EEG was recorded using a 64-channel active electrode ActiCap (BrainProducts) system (500 Hz sampling rate). EEG was filtered online between 0.016Hz and 125 Hz, and referenced to the left mastoid.

During EEG capping, participants filled out the VVIQ questionnaire (see Participants) and were familiarised with the imagery scenes. Once all impedances were brought down < 25 kΩ EEG recording was started and participants began practicing the imagery task, first without, then with intervening stimuli (18 trials each).

Stimuli were shown on a DELL S2421HGF 07 LED screen (23.8", 1920 x 1080 px, 120 Hz refresh rate). Stimulus presentation was controlled using Matlab and Psychtoolbox (Brainard, 1997).

EEG preprocessing

EEG data was preprocessed offline using fieldtrip (Oostenveld et al., 2011) and custom Matlab scripts. Preprocessing included 1) manual selection and removal of noisy channels, 2) filtering (bandpass-filtering between 0.05 – 100 Hz, notchfilter at 50 Hz and its harmonics to remove line noise), 3) re-referencing to an average reference and 4) independent component analysis (ICA) to remove eye artifacts using the runica algorithm.

The continuous data was then epoched between -200 and 400 ms around the onset of imagery or intervening stimuli. Each epoch was baseline corrected, using the first 200 ms prior to stimulus onset as a baseline, and downsampled to 250 Hz. On average, this resulted in 1237.18 (sd 27.00) epochs for the intervening stimuli, as the exact number of trials with or without intervening stimuli varied slightly across participants.

Decoding

All decoding analyses focused on the intervening images, as a function of the images held in mind during the trial. To investigate the timecourse of processing the intervening stimuli, linear support vector machines (SVMs) were trained to decode either their position or category (see Figure 2), using the CoSMoMVPA toolbox (Oosterhof et al., 2016).

To test whether the task-irrelevant objects were indeed processed, all matching and mismatching imagery trials were first combined, to establish above-chance decoding of their location and category. Then decoding accuracy was compared for intervening objects matching and mismatching imagery at different levels (see Figure 1A): 1) generally matching or mismatching at the domain level (including all categories), 2) matching or mismatching at the category level (including only the cued and uncued category that appeared at the start of the trial) and 3) for the other categories matching or mismatching only at the domain-level (including only those categories that had not been shown as imagery scenes in the current trial).

Classifiers were trained separately for each participant and timepoint, training and testing on single-trial data from all 19 posterior and posterior-occipital electrodes (O1, O2, Oz, P1, P2, P3, P4, P5, P6, P7, P8, PO3, PO4, PO7, PO8, PO9, PO10, POz, Pz).

The epoched data was randomly split for ten-fold cross-validation. In order to maximise the available training data, the training folds contained trials from all conditions of interest, but each testing fold contained only data from a single condition. For the first (domain-general) analysis, classifiers were trained on all domain-matching and mismatching trials combined, but tested individually on either the matching or mismatching trials. For the other analyses, the data from matching/mismatching category and other categories from both domains were combined in the training data. There were more trials in which the intervening image matched only at the domain-level but not the particular category. To create a balanced training set, a random subset of trials including either categories of the matching or mismatching domain was taken, to match the number of trials in the category-match condition. Within each fold, 20% of trials were used for testing.

Decoding was performed separately for each pair of conditions (e.g., position 1 vs. position 2, position 1 vs. position 4 etc.), and those pairwise accuracies were then averaged for each participant and timepoint. Decoding accuracies were compared on a timepoint-by-timepoint basis, as well as averaged within two selected time-windows of equal length (50-200 and 200 – 350 ms). This provided a more sensitive and focused measure for differences within those two time-windows, reflecting either an early, mostly bottom-up driven stimulus response and later categorical processing, including influence of feedback processes respectively.

Statistical tests

For behavioural and time-window analyses, bootstrapped t-tests (10000 iterations, $\alpha = .05$, two-sided) were used to test for differences between conditions. For each iteration, difference scores between conditions for individual participants were sampled with replacement. This allowed us to establish a bootstrap distribution of condition differences and their confidence intervals, without prior assumptions regarding their statistical distributions.

To test for decoding above chance at each timepoint, threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009) combined with permutation tests were used to control for multiple comparisons. TFCE boosts belief in neighbouring timepoints containing signal, representing cluster-like spatial support. This procedure was applied to both the real decoding data and synthetic null data. The final p-value reflects how likely a given TFCE score is, given the maximum TFCE scores across the null-data and thereby accounts for multiple comparisons.

Results

Behaviour

Participants were actively engaging in vivid mental imagery during the task, even in the presence of intervening images, rating their average trial vividness as 3.52/5 (sd 0.69). Participants' trial vividness ratings correlated with their offline imagery vividness scores, as measured by the VVIQ ($r = .48$, $p < .01$). Across categories, animals were imagined more vividly than buildings (CI = [0.21, 0.38], $p < .0001$). Vividness ratings were comparable for trials with intervening images and those without (CI = [-0.01, 0.11], $p = .11$), suggesting that the intervening images did not strongly impair imagery.

In addition, participants were able to distinguish the imagined scene from highly similar foils in the snippet task, again suggesting a vivid and detailed mental image. Average accuracy on the snippet task was 79.19% (sd 6.27). Similar to the mean vividness ratings, performance was better for animals compared to buildings (CI = [2.04, 8.87], $p < .0001$), but was not significantly worse on trials with intervening images compared to those without (CI = [-2.10, 3.37], $p = .64$).

Given that participants imagined the specific categories during the delay period, we could now test whether imagery modulated the representation of the intervening objects shown during the imagery delay.

General decoding of stimulus properties

Before testing how imagery modulated the representations of the intervening objects, we first ensured that the properties of the rapidly appearing and task-irrelevant intervening objects were reliably represented in their evoked responses, decoding their position and category independent of their relation to the imagined scenes.

Position decoding

Position decoding is a measure of early stimulus processing, where retinal location is one of the earliest stimulus properties reflected in the neural response. It is also a measure of later spatial attention directed towards the stimulus. As expected, the intervening objects' position around the central imagery square (upper left, upper right, lower left, lower right quadrant) was represented in the neural response early on, starting from 68 ms after stimulus onset, and was sustained beyond 400 ms (see Figure 2).

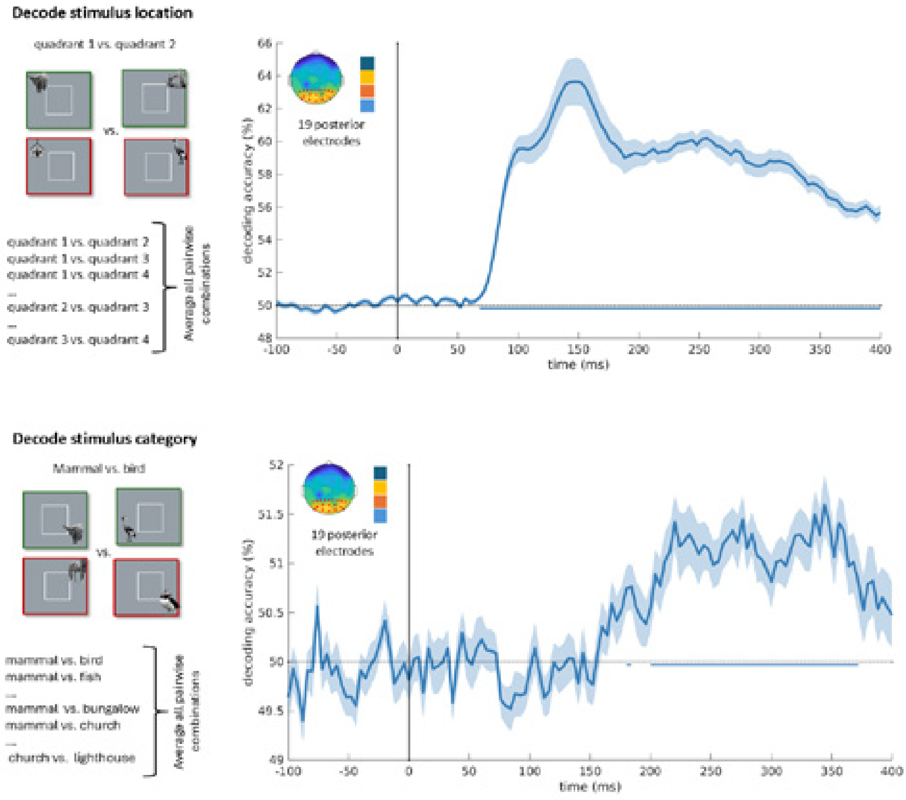

Figure 2: Decoding of intervening stimuli. Position (upper panel) and category (lower panel) decoding, independent of imagery. Coloured lines show significant timepoints for decoding after threshold free cluster enhancement (TFCE, p <.05). Error bars represent standard error of the mean (SEM).

Category decoding

Category-specific representations emerged from 200 ms after onset of the intervening image, and sustained until 372 ms (see Figure 2). This shows that the task-irrelevant intervening images were indeed reliably processed up to the category-level.

Testing domain-general imagery competition

Having established reliable neural responses to the intervening objects, we turned to investigate how perceptual processing of an object was modulated by concurrent mental imagery. First, we tested for competition at the domain level, i.e., comparing the representation of an intervening building when either a building (matching domain) or an animal (mismatching domain) was imagined, and vice versa for animals. Importantly, this comparison was based on visually identical intervening images for both conditions. As imagery scenes from both domains were shown at the beginning of each trial, semantic or low-level priming between domains was also equated. The only difference between matching and mismatching domain images were the specific imagery scenes held in mind.

We tested for differences in decoding accuracy between matching and mismatching domain imagery on a timepoint-by-timepoint basis. In addition, we conducted more sensitive analyses, averaging decoding accuracy within two separate time windows of equal length: an early time window (50-200 ms), broadly encompassing the early visually evoked response and feedforward processing along the ventral visual stream (H. Liu et al., 2009; Seeliger et al., 2018), and a later time window (200-350 ms) when categorical representations of the intervening stimuli emerge and in which feedback-processing might resolve competition (Chelazzi et al., 1998).

Position decoding

The intervening object's position was decodable well above chance, starting from around 76 ms (matching domain) or 68 ms (mismatching domain) and extending beyond 400 ms (see Figure 3). In a timepoint-by-timepoint analysis, no significant differences between conditions survived multiple comparison correction. Focusing specifically on the two time-windows, decoding accuracy was significantly lower for the matching compared to mismatching domain objects in both the early (CI = [0.09, 0.72], p = .01) and late time window (CI = [0.04, 0.77], p = .03). These results provide evidence for similarity-based competition between perceived and imagined objects.

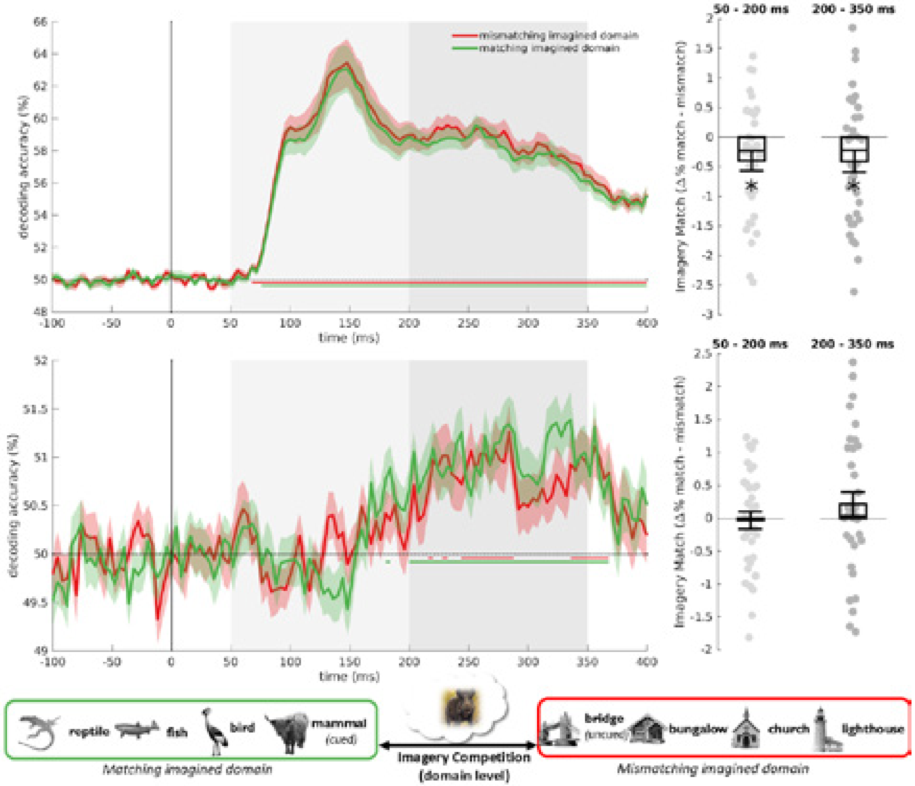

Figure 3: Domain-general competition. Position (upper panel) and category (lower panel) decoding, depending on the domain-level match between presented and imagined objects. Coloured lines show significant timepoints (p < .05) for decoding after threshold free cluster enhancement (TFCE). Shaded squares show early and late time windows for analyses. Right panels show imagery competition (mismatching - matching) averaged across the time points within early and late time windows. Gray dots are individual participants. Error bars represent the standard error of the mean (SEM). *, p < .05.

Category decoding

When the domain of the intervening object matched the domain of the imagined object, category decoding was sustained between 200-368 ms. When it did not match, category decoding was significant across two timepoint-clusters, between 228 – 232 and 336 – 368 ms (see Figure 4). There was no evidence for domain-general competition, neither for individual timepoints or the early (CI = [-0.21, 0.29], p = 0.79) or late (CI = [-0.60, 0.16], p = .25) time windows. This was the same across both low and high vividness trials (early time window: CI = [-0.60, 0.10], p = .18; late time window: CI = [-0.70, 0.32], p = .51).

Imagery competition at the category-level

For a finer-grained analysis, we next tested for imagery competition at the specific category-level, investigating, for example, whether imagining a mammal impaired perception of another mammal. To again equate visual and semantic priming, we compared decoding of the matching category to the other category that had been presented as the uncued imagery scene at the beginning of the trial. This mismatching category was always from the opposite domain as the imagined category.

Position decoding

The position of the current intervening object could again be decoded well above chance from 76 ms onwards, for both conditions (see Figure 4). Between 112 and 132 ms after stimulus onset, we observed an effect of imagery-competition: position decoding for an object belonging to the matching (cued) category was worse than for an object belonging to the mismatching (uncued) category.

This category-specific competition was also apparent in the early time window. Location decoding was worse for matching than mismatching category objects (CI = [0.17, -0.42], $p < .001$). In the late time window, although going in the same direction, there was no significant effect of category-level competition (CI = [-0.91, 0.13], $p = .14$).

Category decoding

The category of an intervening object matching the imagined category could not be consistently decoded above chance. In contrast, category decoding was sustained between 200 - 328 and 340 – 348 ms when the intervening object belonged to the uncued (mismatching) category (see Figure 4). There were no significant timepoint-wise differences, but category-specific competition was significant in both the early (CI = [0.17, 0.80], $p = .001$) and late time windows (CI = [0.00, 0.79], $p = .05$). Therefore, not only extracting the location but also category-information was impaired for intervening objects matching the imagined category.

Domain-level competition for other categories

The previous analyses showed that imagery impaired the representation of the specific imagined category, with a weaker effect at the domain level. This raises the possibility that the domain-level competition we observed was specific to the category-level match condition, which was also included in the domain-level analysis. Therefore, we next tested for competition at the domain level when excluding the imagined category. Again, all semantic and visual priming between domains was controlled for by the retrocue design, and none of the specific categories had been shown as imagery scenes on the relevant trials.

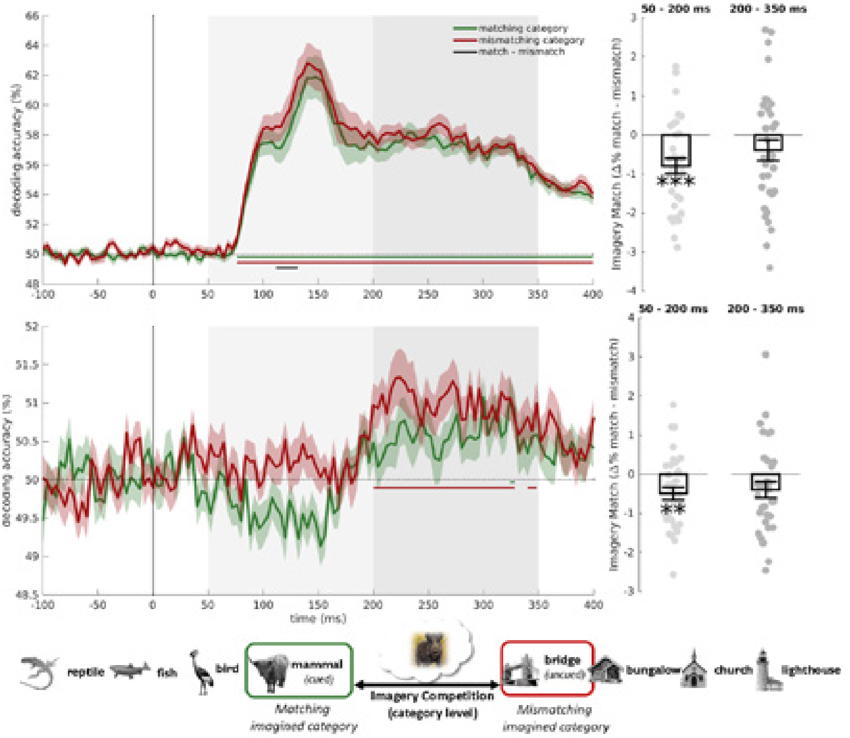

Figure 4: Category-level competition. Position (upper panel) and category (lower panel) decoding, depending on the category-level match between presented and imagined objects. Coloured lines show significant timepoints ($p < 0.05$) for decoding after threshold free cluster enhancement (TFCE). Shaded squares show early and late time windows for analyses. Right panels show imagery competition (mismatching - matching) averaged across the time points of early and late time windows. Gray dots are individual participants. Error bars represent the standard error of the mean (SEM). *, $p < .01$, **, $p < .001$.

Position decoding

Position decoding was significant from around 76 ms onwards, and was sustained until after 400 ms for both conditions (see Figure 5). There was no evidence for domain-level competition across individual timepoints, nor in the early (CI = [-0.28, 0.49], p = .57) or late time window (CI = [-0.04, 0.80], p = .08).

Category decoding

For domain-matching stimuli, category could be decoded between 212-280 and 288-400 ms (see Figure 5). For domain-mismatching stimuli, category information was only present between 312 – 352 ms. Similar to the position decoding results, there was no evidence for domain-level competition across individual timepoints, nor in the early (CI = [-0.42, 0.12], p = .28) or late time window (CI = [-0.51, 0.04], p = .09).

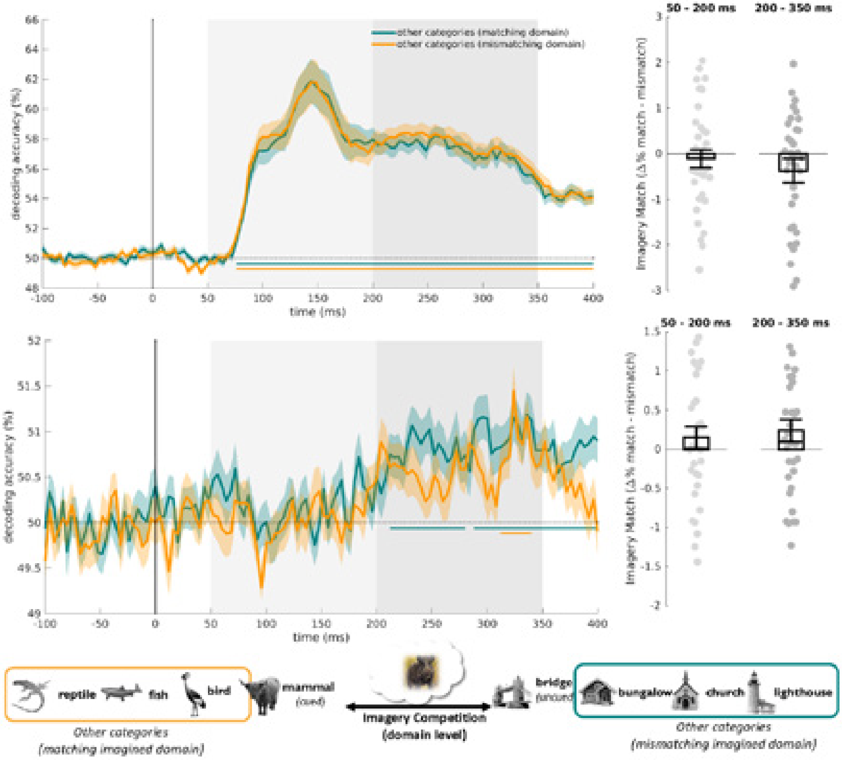

Figure 5: Domain-level competition for other categories. Position (upper panel) and category (lower panel) decoding, depending on domain-level match, excluding the presented categories. Coloured lines show significant timepoints (p < .05) for decoding after threshold free cluster enhancement (TFCE). Shaded squares show early and late time windows for analyses. Right panels show imagery competition (mismatching - matching) averaged across the time points of the early and late time windows. Gray dots are individual participants. Error bars represent the standard error of the mean (SEM).

Comparison between category- and domain-level competition

Was imagery competition indeed specific to the category-level? As the training data for both previous analyses was shared and relied on the same number of trials, we were able to directly compare category and domain-level competition within both time windows (Figure 6).

Position decoding

Category-level competition was higher than domain-level competition within the early time window (CI = [-1.23, -0.14], p = .02). In the late time window, in which neither category- nor domain-level competition were significant, there was also no difference between the two levels (CI = [-0.72, 0.67], p = .92).

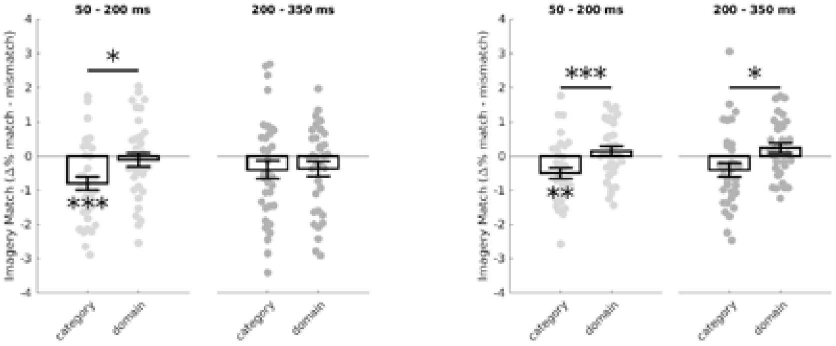

Figure 6: Comparison of category- and domain-level competition within early and late time windows. The left panel shows results for position-decoding, the right for category-decoding. Gray dots are individual participants. All error bars are SEM. *, p < .05; **, p < .01; ***, p < .001.

Category decoding

Similar to the position-decoding results, category-level competition was stronger than domain-level competition in the early time window (CI = [-0.945, -0.32], p < .0001). The same was true in the late time window (CI = [-1.08, -01.15], p = .01).

Discussion

In the current study, we tested a possible consequence of the neural overlap between imagery and perception (Dijkstra et al., 2019; Kosslyn et al., 2001; Pearson, 2019), investigating whether concurrent imagery impairs the processing and representation of task-irrelevant objects. We hypothesised that imagery competes with and degrades the representation of objects evoking similar neural representations as the currently imagined stimulus, i.e., when they were from the same domain or category. To create vivid and detailed visual imagery, recruiting visual cortex, participants were cued to imagine a scene shown at the beginning of the trial as detailed as possible, and had to distinguish between cutouts (snippets) of the imagined and highly similar foil scenes at the end of the trial. The design of the study allowed us to densely sample (eight objects per trial) the responses to intervening visual stimuli during ongoing imagery, comparing evoked responses to visually identical intervening visual stimuli as a function of their similarity with the imagined objects, and separately assess competition at the category- and domain-level.

Participants were able to create vivid and detailed mental images, which they could distinguish from nearly identical foils. Despite this focus on the imagery task, the intervening stimuli evoked well-discriminable responses, allowing us to decode both their position and specific category. We presented two stimuli from different domains and categories at the start of each trial, with a later retrocue indicating which had to be imagined, thereby equating any visual or semantic priming. In line with our hypotheses, our time window analyses showed evidence for imagery-related interference (reflected in worse multivariate decoding) when imagery and perceived objects shared the same category. Both position and category decoding were impaired for intervening objects matching the imagined object at the category level.

Imagery interference was specific to the category-level, with no evidence for broader domain-level competition if the imagined object belonged to a different category within the same domain. Thus, imagining a mammal (e.g., a wolf) impeded perceptual processing of another mammal (e.g., a cow), but not of another animal category (e.g., a reptile). The distinction between category and domain-level interference can be explained by the varying degree of representational similarity between objects that share the same category or merely the same domain, given considerable within-domain variability between representations of different categories, especially for animals (Connolly et al., 2012; Sha et al., 2015; Thorat et al., 2019). In addition, the participants' task, holding in mind a precise image that needed to be distinguished from a highly similar foil, likely resulted in encoding the image in terms of detailed visual features (some of which may be shared within categories), but did not emphasise the membership to a particular domain. Thus, our results generally mirror findings demonstrating similarity-based competition between the representations of multiple external objects in higher-level and object-selective visual cortex (Bao & Tsao, 2018; Cohen et al., 2014; Kliger & Yovel, 2020; Yeh et al., 2024).

At which level of processing do imagery and perception compete? Using EEG allowed us to directly test *when* imagery modulated perceptual processing. We found evidence for competition already at relatively early processing stages: imagery-competition effects were generally found in the 50-200 ms time window and category-level match also interfered with position-decoding specifically around 110-130 ms. Given the onset of this interference with respect to the overall temporal profile of position and category-decoding, it is likely that imagery activated the representation of category-diagnostic mid-level features, impairing processing of matching perceptual input, already before stable category-representations of the intervening images emerged. While the overlap in visual features between imagined and intervening objects was only moderate, certain diagnostic features were shared between exemplars of the same category. That imagery content affected also position-decoding may seem surprising, however there is evidence that object-representations within the ventral stream show substantial spatial specificity (DiCarlo & Maunsell, 2003; Hong et al., 2016; Schwarzlose et al., 2008). Therefore, position decoding in the time range in which we found imagery interference, is likely not agnostic to feature information, but could instead capitalise on the spatial specificity of these object representations. In general, early competition by imagery is in line with previous evidence showing that early perceptual representations can be re-activated by imagery (Dijkstra et al., 2020), but noteworthy given the only limited visual overlap between imagined and perceived objects. The specific time range of imagery-interference effects may, however, depend on the task and imagery-detail. In our task, a detailed visual image had to be re-activated after a retrocue, encouraging re-activation of simple visual features and early visual regions. It is a relevant open question whether similar early interference effects can also be found for objects that have to be imagined in less detail or which are not visually cued in this way, but, e.g., through a word cue.

The current results provide evidence that top-down driven representations during imagery can interfere with the bottom-up driven representation of external objects, akin to the similarity-based competition between multiple external objects. There may, however, be some relevant differences compared to competitive interactions between multiple external objects. First, visual activity evoked by imagery is typically much weaker than perception (Koenig-Robert & Pearson, 2021; S. H. Lee et al., 2012; Pearson, 2019). Further, the degree of overlap between imagery and perception varies not only across the visual hierarchy, but also across different cortical layers (Bergmann et al., 2024; Iamshchinina, Kaiser, et al., 2021) and frequency bands (Stecher & Kaiser, 2024; Xie et al., 2020), indicating overall incomplete overlap. Thus, holding in mind another object may be less detrimental to perception compared to it being actually present in our visual field, in line with overall robust decoding of stimulus properties for the intervening objects and numerically small competition effects.

Many previous studies, including other chapters of this thesis, have demonstrated or argued that top-down activation of visual representations (either imagined or held in working memory) facilitates perception and enhances neural responses for matching input, sometimes with superficially similar retrocue designs (Cochrane et al., 2021; Gayet et al., 2013, 2017; Ishai & Sagi, 1987; Keogh & Pearson, 2011; Pearson et al., 2008; Sulfaro et al., 2024). Here we provide neural evidence for the reverse effect, showing that attending to internal representations does not always facilitate – and can even impair – perception of similar external objects. The time course of decoding for matching and mismatching stimuli also rules out that matching stimuli were suppressed in a reactive manner following initial attentional capture (Gaspelin & Luck, 2018; Moher & Egeth, 2012). This raises the obvious question under which conditions internal and external representations either compete or prime one another. A crucial factor is likely the relationship between imagined and perceived objects, both with respect to whether their interaction is necessary, and the similarity between their representations. Facilitation typically arises in tasks in which the internal representation reflects a task-relevant external object and needs to be matched to incoming (external) input, e.g., in the form of an attentional template. These facilitation effects are at least reduced when external objects do not reflect likely targets, do not have to be responded to, or reflect likely distractors (Carlisle & Woodman, 2011; Olivers & Eimer, 2011). Importantly, we find not only a lack of facilitation, but instead evidence for worse representation of similar objects, suggesting that concurrent imagery reduces the neural resources for representing similar external objects, as both representations should remain separate from each other. In contrast to typical working memory or search tasks, our imagery task encouraged to keep the representation of imagined and external objects fully apart. The intervening objects were entirely task-irrelevant, reflecting distinct exemplars and were not predicted by imagery, nor could they be used to refresh participant's memory of the imagined object. Such objects lying nearby in feature space without overlapping fully have also been found to lead to strongest competition (Kiyonaga & Egner, 2016), but in many tasks finding facilitation, the internal and external representations overlap fully. These task differences likely modulated the interactions between internal and external stimuli, resulting in the observed interference.

Our findings open important questions and avenues for future research. First, varying similarity in a broader and more continuous manner (including a fully matching intervening image), would allow to test the overall profile of facilitation and competition as a function of similarity. In addition, the representational similarity between stimuli across the visual stream could be established using fMRI, allowing to further investigate at which level of the visual hierarchy competition occurs. Using fMRI instead of EEG would make it easier to measure both the strength of the perceived and imagined (which is likely less temporally precise and smeared out; Shatek et al., 2019) representations simultaneously, allowing to test for a direct trade-off between the two.

In conclusion, we find neural evidence that imagery can impair perceptual processing and the representation of task-irrelevant objects of the same category. Our results shed light on the interactions between internally and externally generated representations co-existing within visual cortex.

Chapter 5

General Discussion

Internally-driven (mental) representations of relevant objects, which can be reflected in top-down activity in visual cortex, are thought to support efficient and goal-directed behaviour. As attentional templates, some of these internal representations are thought to bias processing in favour of target-matching visual input. Over the past decades, we have learnt much about their neural basis and content. Yet, there are still important open questions about how these internal representations can support real-world behaviour. Are they context-dependent? Do these preparatory biases indeed support attentional guidance, and do they necessarily reflect the cued target, or rather the most informative features for search? Do internal representations also sometimes compete with the processing of external objects? In this thesis, I have investigated the dynamic interplay between internal representations in visual cortex and the external world - how internal representations are shaped by regularities in the external world, and how these internal representations in turn shape what we perceive.

In Chapter 2, we tested the flexibility of the attentional template for complex two-dimensional target shapes using a series of behavioural experiments. Participants were cued to search for shape targets within two probabilistic contexts, in which the target shape was unique and distinct from distractors in one feature dimensions. Across multiple experiments, we showed that different dimensions were emphasised in the template, depending on their diagnosticity in the current context. This was reliably found in blocked contexts, in which the diagnostic dimension repeated across trials. In addition, we tested whether the template could also be adapted based on trial-by-trial context cues. We found that this was the case, but only when the two distractor contexts were spatially separate, in addition to different background colour cues. These results support the emerging view that the attentional template is not a static and veridical representation of target features, but a representation that can be flexibly adapted for efficient search. Importantly, our results show that we can indeed flexibly emphasise distinct dimensions of a unitary shape, and that this can be done both based on recent selection history as well as trialwise context cues.

In Chapter 3, we investigated preparatory neural activity (the putative neural basis of the attentional template) in context-guided search, using fMRI and concurrent eye tracking. Participants searched for target objects that were associated with different attention-guiding “anchor” objects across two contexts. We found that preparatory activity reflected the relevant guiding object in the current context, but there was no evidence for a preparatory target representation. By dissociating between features that were relevant for attentional guidance and features that were relevant for target identification, we showed that preparatory activity primarily supports the former, reflecting a flexible (context-dependent) and internally-generated guiding template rather than the features of the cued target. The preparatory template can thus not only reflect a somewhat distorted version of the target, but can also represent entirely different objects instead. This provides evidence for the hypothesised distinction between target and guiding templates and sheds light on the functional role of preparatory activity in context-guided search.

In Chapter 4, we moved to mental imagery and EEG to test how internal and external signals co-exist and compete in visual cortex. We found that imagining an object of the same category impaired perceptual processing of external objects, as compared to objects of a different category. This suggests content-specific competition between representations of imagined and perceived objects. This competition was specific to same-category objects, such that it was absent for objects that were within the same domain (animals or buildings) but from a different category (e.g., seeing a mammal and imagining a bird), likely because these representations were less similar. While this thesis, and research overall, has largely focused on the facilitatory effect of internal representations, especially during visual search, here we also provide neural evidence for the opposite effect (i.e., interference) in an imagery task. This study provides new insights into the interactions of top-down and bottom-up signals within visual cortex.

In this final chapter, I will now discuss and integrate these findings, in the light of current theories of attention and competition as well as real-world search.

Preparatory templates for real-world attentional guidance

For efficient attentional selection, an internal representation of the target, the attentional template, is traditionally thought to support both attentional guidance and target identification. However, the conceptualization of the attentional template as a unitary, veridical target representation has been recently questioned (Wolfe, 2020, 2021; X. Yu, Zhou, et al., 2023). There is growing, mostly behavioural, evidence arguing for a dissociation between guiding and target representations. Can these also be dissociated at the neural level? The idea that top-down preactivation of target-selective neurons form the neural basis of such attentional templates is an important cornerstone of current attention theories (Buschman

& Kastner, 2015; Carrasco, 2011; Desimone & Duncan, 1995). Despite growing experimental evidence for the existence of these preparatory biases, however, it was not clear whether they would indeed support guidance in particular or whether they could reflect anything other than veridical features of the cued target category. In Chapter 3, we explicitly dissociated between information that was relevant for guidance (in the form of associated anchor objects) and target identification. We found that preparatory activity in LOC, as well as IPS, supported context-guided search and reflected an internally-generated guiding template, rather than necessarily veridical target features.

What is the content and format of this preparatory guiding template in real-world search? Guidance processes should be optimised to operate fast, at the potential cost of accuracy (X. Yu, Zhou, et al., 2023). Attentional guidance has mostly been linked to relatively low-level features (Wolfe & Horowitz, 2004, 2017), as has been feature-based attention (Maunsell & Treue, 2006), a neural mechanism allowing for global feature-based attentional guidance. From this perspective, one might anticipate a guiding template to reside mostly in these early visual regions. Interestingly, however, we found evidence for a guiding template in object-selective LOC. This suggests that the template was encoded at least in terms of mid-level shape features, e.g., the width of the table surfaces, rather than, e.g., simple luminance differences (looking for a dark or bright table independent of its shape), that would be encoded in early visual cortex. While perhaps surprising especially when comparing it to the results of more abstract, lab-based search tasks, our findings are more in line with previous research studying efficient attentional allocation for real-world scenes and objects. These have shown that preparatory templates for real-world objects in naturalistic scenes are typically encoded within object-selective regions of the ventral visual stream (Gayet & Peelen, 2022; Peelen & Kastner, 2011; Soon et al., 2013). In addition, recent studies found spatially global gain-enhancement (the neural basis of feature-based guidance) not only for simple features, but also for at least some categories such as faces and human bodies (Störmer et al., 2019; Thorat & Peelen, 2022). Category-based attention modulated the response to such target objects early on during stimulus processing (Kaiser et al., 2016). Overall, this shows that efficient and early modulation of attention is possible at the level of object categories. A guiding template defined in terms of at least mid-level, often category-diagnostic features, would allow for more invariance to changes in viewpoint, postures or lighting compared to more low-level templates (Peelen & Kastner, 2011; Reeder & Peelen, 2013). Yet, this does not mean that an efficient template must be abstract enough to match all instances of a given category. In our context-guided search task (Chapter 3) the preparatory template created by participants was likely not highly detailed, but precise enough to distinguish between two exemplars of the same category that were easily visually distinguishable. Overall, even when a low-level template might in principle be sufficient in a particular situation or context, the pre-activation of a mid- or higher-level template may arise quite automatically. This is also in line with our results from Chapter 2, in which overall good performance on context-mismatching trials suggested participants still encoded a more integrated representations including both dimensions of the target shape, in which one feature-dimension was emphasised, but was not relied on exclusively.

Our findings are also further confirmed by results of another very recent study (Zhou & Geng, 2025). In this study, probabilistic associations between scene images and faces could be used for attentional guidance. Following the presentation of a cued target face, preparatory activity in scene-selective parahippocampal place area (PPA), as well as a part of the inferior frontal junction (IFJ), reflected the associated attention-guiding scene. This again indicates that preparatory activity reflects a guiding template encoded in category- or object-selective regions that can include associated objects and is defined in terms of mid- or higher-level visual features.

Overall, preparatory biases in visual cortex should therefore be well suited to support attentional guidance towards real-world objects in structured scenes, playing a major contribution to efficient attentional selection in naturalistic search.

Is the preparatory guiding template sensory-like?

Within this thesis, I have emphasised the sensory-like nature of preparatory activity, as previous studies have found overlap between visually-evoked responses and preparatory activity (Chelazzi et al., 1993; Gayet & Peelen, 2022; Peelen & Kastner, 2011; Stokes, Thompson, Nobre, et al., 2009). Surprisingly, while we found that information about the guiding object (table) was robustly encoded in LOC in Chapter 3, we did not observe cross-decoding between preparatory activity for the anchor tables and visually evoked responses to the same objects. A lack of generalisation from preparatory activity to sensory-evoked responses has also been found in some other studies (Chen et al., 2024; Gong et al., 2022; Mendoza-Halliday et al., 2024). Before interpreting our results as evidence for non-sensory templates, two aspects have to be considered. First, such absence of evidence is hard to interpret and does not equal evidence of absence. More sensitive analyses could thus still uncover a sensory-like code. Second, as just mentioned, given that this template was encoded in LOC, even though shape representations in this area may not be exclusively visual (Amedi et al., 2007; Peelen et al., 2014), this could still suggest that the preparatory template is encoded in terms of visual features, specifically including mid- or high-level shape features rather than a purely abstract, non-sensory representation.

That being said, specifically guiding representations may potentially be less similar to visually evoked responses as captured in a 1-back task, which requires identification of the target. This task likely emphasises more detailed features, whereas a guiding representation should include primarily coarser features that are easy to distinguish in the periphery. Does the representation of a target object change once identification is required? In a recent study, it was argued that during a delay period, attentional templates may be encoded in a default non-sensory state, that was decodable from sensory regions but did not overlap with visually evoked responses (Chen et al., 2024). However, following the presentation of an (even unrelated) stimulus in the delay period, acting as a “ping” to probe a latent representation (D. H. Duncan et al., 2023; Rademaker & Serences, 2017; Wolff et al., 2017), a sensory-like template emerged. This could suggest that a detailed, sensory-like code is especially useful and quickly instantiated at the moment incoming visual input needs to be compared to the target, while it is less useful to sustain this detailed representation across the whole delay period.

Flexibility and stability of the attentional template(s)

An adaptive system should attend to those objects and features that are the most informative and likely to reduce uncertainty, e.g., related to the presence of a relevant target object (Gottlieb et al., 2013; Gottlieb & Oudeyer, 2018). In a dynamic world, however, informative or diagnostic features can change with context. How are attentional templates adjusted to these changing environments? There is now growing evidence that attentional templates are not static or necessarily veridical object representations, but instead are biased towards informative and relevant features (X. Yu, Zhou, et al., 2023). Our results in Chapter 2 and 3 add to this evidence, by addressing open questions about the specific (neural) mechanisms behind this flexibility and their context-dependency.

As briefly reviewed in the introduction, regularities in our environment can be learnt and impact attentional selection through different mechanisms and at different timescales. In Chapter 2, we first observed adaptation based on recent search and selection history in emphasizing diagnostic dimensions of the target in blocked contexts. Importantly, we found that learning and using those block-based regularities was fast and appeared non-strategic. Participants who reported having no explicit awareness of the dimension manipulation show reliable adaptation to context. For extended searches in a stable environment, such selection-history based biases can lead to powerful attentional biases and drive highly adaptive and effortless changes in the template (Kristjánsson, 2022). Such trial-based regularities are also the most commonly studied when investigating how the attentional template is adapted to the current context.

However, not every search is as extended, and some regularities are highly consistent within a given environmental context, but variable across contexts. For instance, trains in the Netherlands are often yellow, red/white in Germany, and another shade of red in France. In order to flexibly adapt our attentional templates to these regularities, selection history alone would not suffice, as these biases would not be context-dependent. Can the content of attentional templates be truly contextualised, i.e., do we use context cues to learn and use different diagnostic information or features in different environments? To test this, both in Chapter 2 and 3, we used trial-by-trial context cues to indicate diagnostic dimensions or object associations in the current context. In both studies, these cues were successfully used. Participants' attentional templates were flexibly adjusted based on different contextual regularities: either based on shared vs. distinct feature dimensions of the target and distractors, or object associations. The regularities in both studies were relatively complex, either pertaining to diagnosticity of whole feature dimensions, or object-associations that reversed across contexts, as compared with a more simple change in feature/object probabilities. Despite this complexity, we found clear evidence for context-dependent attentional templates across both studies. Importantly, this flexibility was not only shown in behavioural measures, but also for preparatory biases in visual cortex in Chapter 3. This flexible attentional allocation is likely an important factor for efficient attentional selection in complex but structured environments.

Nonetheless, contextual cues and associations may not always be used. Given that diagnostic features change probabilistically across contexts, the visual system likely faces a trade-off between adjusting an attentional template to the current context, which may be potentially effortful and also less useful when the context quickly changes again, compared to a more general, stable representation that is not optimised to the current context. In the search task in Chapter 2, this would mean always weighting shape dimensions equally, independently of their diagnosticity in the current context, or to focus on individual shape features that were stable across contexts. In Chapter 3, this could mean using (context-independent) target-features for guidance, ignoring the changing context-dependent association with anchor objects. We saw some evidence for this trade-off, especially in Chapter 2 where context-cues were not always used, but only shaped the attentional template when different contexts were not only indicated by different background colours but also space. Also other studies have shown that using contextual associations or trial-wise cues to bias attentional templates is most common in relatively hard search tasks (Bravo & Farid, 2016; Scolari et al., 2012, 2012; Scolari & Serences, 2009; Zhou & Geng, 2024).

Are such context-dependent attentional templates therefore strictly limited to very difficult, effortful searches? Related questions are whether participants would have used those regularities if they had not been explicitly informed about them, or if the associations were less consistent. First, it is probable that context-associations are more likely to be learnt and used when search is otherwise suboptimal. If search is successful with the current template, there may be little use to adjust it. Across all our search tasks, however, performance was overall well above chance, suggesting that contextual associations are still used when a search task is challenging, but not impossible without a context-dependent template. In addition, object associations have been shown to influence search also when associations are not deterministic but only probabilistic (Boettcher et al., 2018; Mack & Eckstein, 2011; Zhou & Geng, 2024), with recent evidence showing also evidence for a change in preparatory activity based on these probabilistic associations (Zhou & Geng, 2025). With respect to learning, there is now evidence that learning of some regularities, e.g., the frequent colour of the target, or value associated with them, can be structured by contexts indicated by different scene backgrounds on which the objects are superimposed (Anderson, 2015; Kershner & Hollingworth, 2022). Those regularities were quickly learnt without explicit instructions to do so and modulated attentional guidance in a context-dependent manner. For more complex regularities, as in our tasks, this learning may take longer but is likely not impossible, especially when stronger context cues are used. This could be seen in Chapter 2, where informing participants about colour-based context-cues itself was not sufficient to evoke context-dependent templates, but stronger context cues including spatial separation of contexts did, likely because those made it easier to form and retrieve relevant contextual associations. An important aspect of real-world search is that naturalistic environments can offer a wealth of multimodal context cues that can be used to structure learning. For instance, before we even step onto the platform at the train station in Germany or the Netherlands and begin to look for our either red or yellow train, we will have picked up on different languages in the announcements and conversations of the people around us, seen different brands of shops in the station hall etc; all of which would help us to form and retrieve a context-dependent memory representation for the train colour. This is unfortunately difficult to recreate in experimental search tasks, but should make the use of context-dependent templates in real-world search quite effortless.

Our findings, as well as previous results, strongly suggest that we can create different attentional templates when searching for the same complex, real-world object in different contexts (Addleman et al., 2024; Bravo & Farid, 2012, 2016; Kershner & Hollingworth, 2022; Lau et al., 2021). In addition, we are equipped with a range of different mechanisms, ranging from brief selection history biases, quick learning of contextual associations mediated by the hippocampus (Schapiro et al., 2012; Solomon et al., 2024) and finally long-term changes in cortical object representations (Kaiser et al., 2019; Kaiser & Peelen, 2018; Sigala & Logothetis, 2002) that should allow for an optimal trade-off between stability and flexibility of attentional allocations for short and long term regularities.

Flexibility of target and guiding representations

The distinction between target and guiding templates raises the question whether both are equally flexible or context-dependent. This division could be another way to optimally balance flexibility and stability, and given their distinct roles, the optimal point in this trade-off may differ for target and guiding templates. The target template should be detailed, closely related to our long term memory representations for the particular target object or category, and will likely be retrieved from memory across varying retrieval contexts (Boettcher et al., 2020; Wolfe, 2020, 2021). As such, it may be possible to quickly adjust the decision criterion for identification, but it is likely that the underlying representations remain relatively stable. A guiding template on the other hand, which does not need to be as precise but rather allow for quick selection in the current context, may be faster to adapt to regularities in the current context. As outlined in the introduction, there is some evidence for different magnitudes of adaptation between identification- and guiding processes for repeating regularities in the same context, with stronger adaptation for early guidance processes. Accordingly, our results in Chapter 3 provide evidence for the flexibility of early attentional processes, such as the content of the preparatory guiding template and the first eye movement of the scene based on context-dependent regularities. To test this further, it may be useful to also explicitly test for a change in speed of adapting to a given context for both guiding and identification processes.

Competition in visual cortex and the role of internally generated representations

Not only during search, but also during other real-world behaviour, externally-driven stimulus representations co-exist and interact with internally-driven top-down signals that target perceptual representations in visual cortex, for instance during mental imagery. I have begun this thesis by discussing competitive interactions between external stimuli, viewed through the lens of biased competition theory (Desimone, 1995). There is abundant experimental evidence that bottom-up sensory input and top-down signals target the same perceptual representations in visual cortex (Battistoni et al., 2017; Dijkstra et al., 2019; Gazzaley & Nobre, 2012). Competition between external and imagined representations, especially if they are similar, should then be a natural consequence of this overlap (Sulfaro et al., 2023), for which there was, however, no clear neural evidence so far. In Chapter 4, we found that these internal signals interfered with processing of intervening images during mental imagery when both imagined and perceived objects matched at the category level. This suggests that internally and externally generated signals indeed compete for representation in visual cortex. Our findings provide initial neural evidence for content-specific competition between internal and external representations during imagery, highlighting an important new facet in their interactions. This opens relevant questions for future research. Given stronger overlap between both types of representations in higher-level visual cortex (S. H. Lee et al., 2012), competition between internal and external representations should be highest in those regions. Intriguingly, however, our findings suggest that concurrent imagery could already impair perceptual processing at a relatively early stage (between 50 – 200 ms), before stable categorical representations emerged. This finding could be followed up by fMRI, testing in which brain regions representations imagined and perceived objects compete. In addition, it is an open question whether attention can resolve this competition, when directed towards either imagined or perceived objects, similar to its effect in resolving competition between external objects.

Can these findings be integrated with the facilitatory role that top-down signals are typically thought to play, especially in the form of attentional templates? A potential explanation is that differences in task demands, e.g., whether internal representations are specifically recruited to resolve competition between external stimuli and interact with ongoing visual stimulation or should be kept apart as in our imagery task, determines how internal and external signals interact. Another likely key difference is that we investigated the effect of concurrent imagery on objects that could be closely related to, but never directly matched, internal representations. For instance, our participant's held in mind a highly specific image of a cow, which interfered with concurrent processing of a horse. Had we actually shown the exact same image of a cow as intervening image, we may have observed evidence for facilitation, as the same neural representation would then be activated by both top-down and bottom-up processes, without competition between both. This could potentially also be observed in search tasks. In their standard form, models of biased competition and feature-based attention would suggest a linear relationship between the match of the internal representation and the amount of top-down enhancement of the neural response (Desimone, 1995; Maunsell & Treue, 2006). There is, however, also experimental evidence that attentional gain enhancement does not follow an entirely linear profile, but shows lowest gain enhancement or suppression for objects or features that lie close-by to attended objects in feature space (Fang & Liu, 2019; Kiyonaga & Egner, 2016; Störmer & Alvarez, 2014; Treue, 2014). This centre-surround profile could sharpen the representation of attended (internal or external) objects, reducing interference by others that lie close-by in feature-space. At this point, the exact profile of either facilitation or competition between internal and external representations is not known yet, and how to integrate and combine these findings raises important new questions for future research.

Limitations

An important goal of this thesis was to investigate the neural and behavioural mechanisms of efficient search within structured environments, with the goal of understanding efficient real-world search. While always motivated by aspects of the real world, e.g., guidance by co-occurring objects or the fact that we typically process concurrent visual input while engaging in imagery, some aspects of real-world experience were not re-created. For instance, we measured search performance for briefly flashed scenes in which participants could not make unlimited eye movements. Some real-world searches, e.g., for missing keys in my living room, can be temporally much more extended and involve long series of eye movements. For such searches, more information about the current environment can be gathered and integrated across subsequent fixations, and other strategic components become additionally relevant, e.g., keeping in mind which part of the search scene has already been searched through and focusing on the still unsearched parts (e.g., through inhibition of return Klein, 2000). Similar to the search time, also the timescale for learning regularities was restricted, focusing on those learnt within the time range of a 1-2 hours experimental session. This showed that novel regularities could be quickly learnt and used, constituting an important skill for efficient real-world behaviour, but we cannot directly speak to the role of long-term learning and plasticity in shaping attentional processes. Another relevant distinction to consider is how the search target or imagery content were defined. An important aspect of real-world attentional templates and imagery are that they are fundamentally self-generated and determined by the participant. For instance, what someone attends to is determined by their intrinsic goals, e.g., we look out for cars and bikes because we want to cross the road, not based on some external cue telling us to search for them. We showed that the content of internal representations can be disentangled from this external cue, but the target objects and to-be-imagined objects were still defined by the experimenters, which may e.g. affect the source regions through which biases in visual cortex are instantiated.

Finally, all participants were young, healthy adults. Visual search, attention and top-down control mechanisms evolve across the lifespan and decline with age (Braver & West, 2011; Madden et al., 2004; Wiegand et al., 2014), and the same may potentially hold true for imagery abilities (Gulyás et al., 2022). Especially relevant for this thesis is the finding that older adults may show decreased statistical learning, but also stronger reliance on long-term regularities or schemas (Cox & Aimola Davies, 2022; Wynn et al., 2020). Thus, we may expect different trade-offs in stability and flexibility of attentional templates across the lifespan, or in the strength of competition between internal and external representations, topics that were not investigated here.

Conclusions and outlook

Think of the last time you stood in a crowded train station, listening to an exciting audiobook while searching for a much-needed cup of coffee and then for the correct train, while also trying not to run into anyone around you. How do our brains efficiently prioritise relevant objects given our rich and complex external, and also internal world? This is a challenging question, perhaps the most important one, that has intrigued cognitive neuroscientists for decades. Through extensive research, we have begun to understand how internal representations in visual cortex can support attentional selection and efficient, goal-directed behaviour.

In the current thesis, I have investigated open questions about their content, flexibility and functional role during real-world behaviour, in which the external environment provides a vast amount of contextual regularities and in which top-down biases have to co-exist and interact with concurrent visual input. In order to tackle these big questions, I have isolated some relevant aspects of our everyday experience, for instance by creating simpler, structured environments that participants had to search through. While this meant simplifying some of the many intricacies of real-world behaviour, this allowed me to test critical assumptions about the nature of our internal representations in a controlled manner. This approach provides important new insights. I started out this thesis with the traditional view of a single, veridical attentional template, reflected in preparatory activity in visual cortex. At the end, we now have evidence for not only one, but rather two templates, with flexible and context-dependent representations in visual cortex that support attentional guidance and may not only facilitate, but also directly compete with external visual input.

The current studies have shed new light on the nature of attentional templates and our internal representations in a broader sense, providing insights what searching humans or internal representations *can* do. For a full understanding of whether and how this set of tools is used in real-world situations, this approach also needs continuous feedback. For future research, I hope that even more naturalistic search paradigms, for instance combining virtual reality (VR) with neuroimaging and eye tracking, can lead us to refine our understanding of search strategies and attentional selection in real world scenes. This could be especially important to further understand how participant's determine what to attend to intrinsically, e.g. based on the current relevance of an object within an extended task, as well as their own curiosity and drive for information (Gottlieb et al., 2013; Gottlieb & Oudeyer, 2018; van Lieshout et al., 2020).

Further understanding the content of our internal representations, and how they interact with incoming visual input will likely remain an important question for the upcoming decades. Next to behavioural paradigms, we now have increasingly refined analysis methods to reconstruct them from neural data, which could allow us to understand variations in their content across time, trials or individuals. In addition, especially related to their competitive interactions in visual cortex, there are many open questions about the mechanisms and consequences of this competition, and how it may be reduced or avoided. There is still much to explore!

References

Addleman, D. A., Rajasingh, R., & Störmer, V. S. (2024). Attention to object categories: Selection history determines the breadth of attentional tuning during real-world object search. *Journal of Experimental Psychology: General*, *153*(6), 1568–1581. https://doi.org/10.1037/xge0001575

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. https://doi.org/10.1016/j.cub.2013.05.065

Aly, M., & Turk-Browne, N. B. (2017). How hippocampal memory shapes, and is shaped by, attention. In *The Hippocampus from Cells to Systems: Structure, Connectivity, and Functional Contributions to Memory and Flexible Cognition* (pp. 369–403). Springer International Publishing. https://doi.org/10.1007/978-3-319-50406-3_12

Amedi, A., Stern, W. M., Camprodon, J. A., Bermpohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, *10*(6), 687–689. https://doi.org/10.1038/nn1912

Anderson, B. A. (2015). Value-driven attentional priority is context specific. *Psychonomic Bulletin and Review*, *22*(1), 750–756. https://doi.org/10.3758/s13423-014-0724-0

Anderson, B. A., Kim, H., Kim, A. J., Liao, M.-R., Mrkonja, L., Clement, A., & Grégoire, L. (2021). The past, present, and future of selection history. *Neuroscience & Biobehavioral Reviews*, *130*, 326–350. https://doi.org/10.1016/j.neubiorev.2021.09.004

Arita, J. T., Carlisle, N. B., & Woodman, G. F. (2012). Templates for rejection: Configuring attention to ignore task-irrelevant features. *Journal of Experimental Psychology. Human Perception and Performance*, *38*(3), 10.1037/a0027885. https://doi.org/10.1037/a0027885

Awh, E., Belopolsky, A. V., & Theeuwes, J. (2012). Top-down versus bottom-up attentional control: A failed theoretical dichotomy. *Trends in Cognitive Sciences*, *16*(8), 437–443. https://doi.org/10.1016/j.tics.2012.06.010

Baeck, A., Wagemans, J., & Beeck, H. P. O. de. (2013). The distributed representation of random and meaningful object pairs in human occipitotemporal cortex: The weighted average as a general rule. *NeuroImage*, *70*, 37–47. https://doi.org/10.1016/j.neuroimage.2012.12.023

Bahle, B., Kershner, A. M., & Hollingworth, A. (2021). Categorical Cuing: Object Categories Structure the Acquisition of Statistical Regularities to Guide Visual Search. *Journal of Experimental Psychology: General*, *150*(12), 2552–2566. https://doi.org/10.1037/xge0001059

Bahle, B., Matsukura, M., & Hollingworth, A. (2018). Contrasting gist-based and template-based guidance during real-world visual search. *Journal of Experimental Psychology. Human Perception and Performance*, *44*(3), 367–386. https://doi.org/10.1037/xhp0000468

Bahmani, Z., Daliri, M. R., Merrikhi, Y., Clark, K., & Noudoost, B. (2018). Working Memory Enhances Cortical Representations via Spatially Specific Coordination of Spike Times. *Neuron*, *97*(4), 967-979.e6. https://doi.org/10.1016/j.neuron.2018.01.012

Bao, P., & Tsao, D. Y. (2018). Representation of multiple objects in macaque category-selective areas. *Nature Communications*, *9*(1), 1–16. https://doi.org/10.1038/s41467-018-04126-7

Barbosa, J., Lozano-Soldevilla, D., & Compte, A. (2021). Pinging the brain with visual impulses reveals electrically active, not activity-silent, working memories. *PLOS Biology*, *19*(10), e3001436. https://doi.org/10.1371/journal.pbio.3001436

Battistoni, E., Kaiser, D., Hickey, C., & Peelen, M. V. (2020). The time course of spatial attention during naturalistic visual search. *Cortex*, *122*, 225–234. https://doi.org/10.1016/j.cortex.2018.11.018

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, *1396*(1), 92–107. https://doi.org/10.1111/nyas.13320

Beck, V. M., Hollingworth, A., & Luck, S. J. (2012). Simultaneous Control of Attention by Multiple Working Memory Representations. *Psychological Science*, *23*(8), 887–898. https://doi.org/10.1177/0956797612439068

Becker, M. W., Hemsteger, S., & Peltier, C. (2016). No templates for rejection: A failure to configure attention to ignore task-irrelevant features attention to ignore task-irrelevant features. *Visual Cognition*, *23*(9–10), 1150–1167. https://doi.org/10.1080/13506285.2016.1149532

Becker, M. W., & Rasmussen, I. P. (2008). Guidance of Attention to Objects and Locations by Long-Term Memory of Natural Scenes. *Journal of Experimental Psychology: Learning Memory and Cognition*, *34*(6), 1325–1338. https://doi.org/10.1037/a0013650

Becker, S. I. (2010a). Testing a postselectional account of across-dimension switch costs. *Psychonomic Bulletin & Review*, *17*(6), 853–861. https://doi.org/10.3758/PBR.17.6.853

Becker, S. I. (2010b). The Role of Target-Distractor Relationships in Guiding Attention and the Eyes in Visual Search. *Journal of Experimental Psychology: General*, *139*(2), 247–265. https://doi.org/10.1037/a0018808

Becker, S. I. (2013). Simply shapely: Relative, not absolute shapes are primed in pop-out search. *Attention, Perception, and Psychophysics*, *75*(5), 845–861. https://doi.org/10.3758/s13414-013-0433-1

Becker, S. I., Folk, C. L., & Remington, R. W. (2010). The Role of Relational Information in Contingent Capture. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(6), 1460–1476. https://doi.org/10.1037/a0020370

Becker, S. I., Folk, C. L., & Remington, R. W. (2013). Attentional Capture Does Not Depend on Feature Similarity, but on Target-Nontarget Relations. *Psychological Science*, *24*(5), 634–647. https://doi.org/10.1177/0956797612458528

Becker, S. I., Grubert, A., & Dux, P. E. (2014). Distinct neural networks for target feature versus dimension changes in visual search, as revealed by EEG and fMRI. *NeuroImage*, *102*(P2), 798–808. https://doi.org/10.1016/j.neuroimage.2014.08.058

Becker, S. I., Hamblin-Frohman, Z., Xia, H., & Qiu, Z. (2023). Tuning to non-veridical features in attention and perceptual decision-making: An EEG study. *Neuropsychologia*, *188*(January), 108634. https://doi.org/10.1016/j.neuropsychologia.2023.108634

Bergmann, J., Petro, L. S., Abbatecola, C., Li, M. S., Morgan, A. T., & Muckli, L. (2024). Cortical depth pro fi les in primary visual cortex for illusory and imaginary experiences. *Nature Communications*, *15*(1002). https://doi.org/10.1038/s41467-024-45065-w

Bettencourt, K. C., & Xu, Y. (2015). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150–157. https://doi.org/10.1038/nn.4174

Bichot, N. P., Heard, M. T., DeGennaro, E. M., & Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron*, *88*(4), 832–844. https://doi.org/10.1016/j.neuron.2015.10.001

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and Serial Neural Mechanisms for Visual Search in Macaque Area V4. *Science*, *308*, 529–535.

Bichot, N. P., Xu, R., Ghadooshahy, A., Williams, M. L., & Desimone, R. (2019). The role of prefrontal cortex in the control of feature attention in area V4. *Nature Communications*, *10*(1), 5727. https://doi.org/10.1038/s41467-019-13761-7

Biederman, I., Glass, A. L., & Stacy, E. W. (1973). Searching for objects in real-world scenes. *Journal of Experimental Psychology*, *97*(1), 22–27. https://doi.org/10.1037/h0033776

Biederman, I., Mezzanotte, R. J., & Rabinowitz, J. C. (1982). Scene Perception: Detecting and Judging Objects Undergoing Relational Violations. *Cognitive Psychology*, *14*, 143–177.

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Võ, M. L. H. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision*, *18*(13), 1–13. https://doi.org/10.1167/18.13.11

Boettcher, S. E. P., & Nobre, A. C. (2024). Going through the motions: Biasing of dynamic attentional templates. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001665

Boettcher, S. E. P., Nobre, A. C., Boettcher, S. E. P., & van Ede, F. (2020). Functional biases in attentional templates from associative memory. *Journal of Vision*, *20*(13), 1–10. https://doi.org/10.1167/jov.20.13.7

Bouwkamp, F. G., de Lange, F. P., & Spaak, E. (2024). Spatial Predictive Context Speeds Up Visual Search by Biasing Local Attentional Competition. *Journal of Cognitive Neuroscience*, 1–15. https://doi.org/10.1162/jocn_a_02254

Bracci, S., Daniels, N., & De Beeck, H. O. (2017). Task context overrules object- And category-related representational content in the human parietal cortex. *Cerebral Cortex*, *27*(1), 310–321. https://doi.org/10.1093/cercor/bhw419

Brady, T. F., & Chun, M. M. (2007). Spatial constraints on learning in visual search: Modeling contextual cuing. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 798–815. https://doi.org/10.1037/0096-1523.33.4.798

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Braver, T. S., & West, R. (2011). Working Memory, Executive Control, and Aging. In *The Handbook of Aging and Cognition*. Routledge. https://doi.org/10.4324/9780203837665.ch7

Bravo, M. J. (2009). The specificity of the search template. *Journal of Vision*, *9*(1), 1–9. https://doi.org/10.1167/9.1.34.Introduction

Bravo, M. J., & Farid, H. (2012). Task demands determine the specificity of the search template. *Attention, Perception, & Psychophysics*, *74*(1), 124–131. https://doi.org/10.3758/s13414-011-0224-5

Bravo, M. J., & Farid, H. (2014). Informative cues can slow search: The cost of matching a specific template. *Attention, Perception, and Psychophysics*, *76*(1), 32–39. https://doi.org/10.3758/s13414-013-0532-z

Bravo, M. J., & Farid, H. (2016). Observers change their target template based on expected context. *Attention, Perception, and Psychophysics*, *78*(3), 829–837. https://doi.org/10.3758/s13414-015-1051-x

Breedlove, J. L., St-Yves, G., Olman, C. A., & Naselaris, T. (2020). Generative Feedback Explains Distinct Brain Activity Codes for Seen and Mental Images. *Current Biology*, *30*(12), 2211-2224.e6. https://doi.org/10.1016/j.cub.2020.04.014

Britton, M. K., & Anderson, B. A. (2019). Specificity and Persistence of Statistical Learning in Distractor Suppression. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(3), 324–334. https://doi.org/10.1037/xhp0000718

Brockmole, J. R., & Henderson, J. M. (2006). Using real-world scenes as contextual cues for search. *Visual Cognition*, *13*(1), 99–108. https://doi.org/10.1080/13506280500165188

Buffalo, E. A., Fries, P., Landman, R., Liang, H., & Desimone, R. (2010). A backward progression of attentional effects in the ventral stream. *Proceedings of the National Academy of Sciences*, *107*(1), 361–365. https://doi.org/10.1073/pnas.0907658106

Buschman, T. J., & Kastner, S. (2015). From Behavior to Neural Dynamics: An Integrated Theory of Attention. *Neuron*, *88*(1), 127–144. https://doi.org/10.1016/j.neuron.2015.09.017

Cabbai, G., Brown, C. R. H., Dance, C., Simner, J., & Forster, S. (2023). Mental imagery and visual attentional templates: A dissociation. *Cortex*, *169*, 259–278. https://doi.org/10.1016/j.cortex.2023.09.014

Cabbai, G., Racey, C., Simner, J., Dance, C., Ward, J., & Forster, S. (2024). Sensory representations in primary visual cortex are not sufficient for subjective imagery. *Current Biology*, *34*(21), p5073-5082.e5. https://doi.org/10.1016/j.cub.2024.09.062

Carlisle, N. B., Arita, J. T., Pardo, D., & Woodman, G. F. (2011). Attentional templates in visual working memory. *Journal of Neuroscience*, *31*(25), 9315–9322. https://doi.org/10.1523/JNEUROSCI.1097-11.2011

Carlisle, N. B., & Woodman, G. F. (2011). Automatic and strategic effects in the guidance of attention by working memory representations. *Acta Psychologica*, *137*(2), 217–225. https://doi.org/10.1016/j.actpsy.2010.06.012

Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research*, *51*(13), 1484–1525. https://doi.org/10.1016/j.visres.2011.04.012

Carrasco, M., Evert, D. L., Chang, I., & Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception & Psychophysics*, *57*(8), 1241–1261. https://doi.org/10.3758/BF03208380

Castelhano, M. S., & Heaven, C. (2010). The relative contribution of scene context and target features to visual search in scenes. *Attention, Perception & Psychophysics*, *72*(5), 1283–1297. https://doi.org/10.3758/APP.72.5.1283

Castelhano, M. S., & Heaven, C. (2011). Scene context influences without scene gist: Eye movements guided by spatial associations in visual search. *Psychonomic Bulletin and Review*, *18*(5), 890–896. https://doi.org/10.3758/s13423-011-0107-8

Castelhano, M. S., & Henderson, J. M. (2007). Initial Scene Representations Facilitate Eye Movement Guidance in Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(4), 753–763. https://doi.org/10.1037/0096-1523.33.4.753

Castelhano, M. S., & Krzyś, K. (2020). Rethinking Space: A Review of Perception, Attention, and Memory in Scene Processing. *Annual Review of Vision Science*, *9*(40), 1–24.

Chang, S., Zhang, X., Cao, Y., Pearson, J., & Meng, M. (2025). Imageless imagery in aphantasia revealed by early visual cortex decoding. *Current Biology*, *0*(0). https://doi.org/10.1016/j.cub.2024.12.012

Chapman, A. F., Chunharas, C., & Störmer, V. S. (2023). Optimal tuning of feature-based attention warps the perception of visual features. *PsyArxiv*, 1–9. https://doi.org/10.31234/osf.io/hbfup

Chawla, D., Rees, G., & Friston, K. J. (1999). The physiological basis of attentional modulation in extrastriate visual areas. *NeuroImage*, *9*(6 PART II), 671–676.

Chelazzi, L., Duncan, J., Miller, E. K., & Desimone, R. (1998). Responses of neurons in inferior temporal cortex during memory-guided visual search. *Journal of Neurophysiology*, *80*(6), 2918–2940. https://doi.org/10.1152/jn.1998.80.6.2918

Chelazzi, L., Libera, C. D., Sani, I., & Santandrea, E. (2011). Neural basis of visual selective attention. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 392–407. https://doi.org/10.1002/wcs.117

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345–347. https://doi.org/10.1038/363345a0

Chen, Y., Liu, T., Jia, K., Theeuwes, J., & Gong, M. (2024). Dual-format attentional template during preparation in human visual cortex. *bioRxiv*. https://doi.org/10.1101/2024.09.12.602176

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2016). Building ensemble representations: How the shape of preceding distractor distributions affects visual search. *Cognition*, *153*, 196–210. https://doi.org/10.1016/j.cognition.2016.04.018

Chetverikov, A., Campana, G., & Kristjánsson, Á. (2017). Rapid learning of visual ensembles. *Journal of Vision*, *17*(2), 1–15. https://doi.org/10.1167/17.2.21

Chetverikov, A., Hansmann-Roth, S., Tanrıkulu, Ö. D., & Kristjánsson, Á. (2019). Feature Distribution Learning (FDL): A New Method for Studying Visual Ensembles Perception with Priming of Attention Shifts. In S. Pollmann (Ed.), *Spatial Learning and Attention Guidance* (Vol. 151, pp. 37–57). Springer US. https://doi.org/10.1007/7657_2019_20

Chota, S., & van der Stigchel, S. (2021). Dynamic and flexible transformation and reallocation of visual working memory representations. *Visual Cognition*, *29*(7), 409–415. https://doi.org/10.1080/13506285.2021.1891168

Christophel, T. B., Cichy, R. M., Hebart, M. N., & Haynes, J. D. (2015). Parietal and early visual cortices encode working memory content across mental transformations. *NeuroImage*, *106*, 198–206. https://doi.org/10.1016/j.neuroimage.2014.11.018

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J. D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, *21*(4), 494–496. https://doi.org/10.1038/s41593-018-0094-4

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111–124. https://doi.org/10.1016/j.tics.2016.12.007

Chun, M. M., Golomb, J. D., & Turk-Browne, N. B. (2011). A Taxonomy of external and internal attention. *Annual Review of Psychology*, *62*, 73–101. https://doi.org/10.1146/annurev.psych.093008.100427

Chun, M. M., & Jiang, Y. (1998). Contextual Cueing: Implicit Learning and Memory of Visual Context Guides Spatial Attention. *Cognitive Psychology*, *36*, 28–71.

Chun, M. M., & Jiang, Y. (1999). Top-down attentional guidance based on implicit learning of visual covariation. *Psychological Science*, *10*(4), 360–365.

Cichy, R. M., Heinzle, J., & Haynes, J. D. (2012). Imagery and perception share cortical representations of content and location. *Cerebral Cortex*, *22*(2), 372–380. https://doi.org/10.1093/cercor/bhr106

Cochrane, B. A., Townsend, B., Lee, R. C., Legere, J. K., Milliken, B., & Shedden, J. M. (2021). Visual imagery influences attentional guidance during visual search: Behavioral and electrophysiological evidence. *Attention, Perception, and Psychophysics*, *83*(1), 58–66. https://doi.org/10.3758/s13414-020-02208-7

Cohen, M. A., Alvarez, G. A., Nakayama, K., & Konkle, T. (2017). Visual search for object categories is predicted by the representational architecture of high-level visual cortex. *J Neurophysiol*, *117*, 388–402. https://doi.org/10.1152/jn.00569.2016.-Visual

Cohen, M. A., Konkle, T., Rhee, J. Y., Nakayama, K., & Alvarez, G. A. (2014). Processing multiple visual objects is limited by overlap in neural channels. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(24), 8955–8960. https://doi.org/10.1073/pnas.1317860111

Connolly, A. C., Swaroop Guntupalli, J., Gors, J., Hanke, M., Halchenko, Y. O., Wu, Y. C., Abdi, H., & Haxby, J. V. (2012). The representation of biological classes in the human brain. *Journal of Neuroscience*, *32*(8), 2608–2618. https://doi.org/10.1523/JNEUROSCI.5547-11.2012

Cox, J. A., & Aimola Davies, A. M. (2022). Age differences in visual statistical learning: Investigating the effects of selective attention and stimulus category. *Psychology and Aging*, *37*(6), 698–714. https://doi.org/10.1037/pag0000697

Craver-Lemley, C., & Arterberry, M. E. (2003). Imagery-induced interference on a visual detection task. *Spatial Vision*, *14*(2), 101–119. https://doi.org/10.1163/156856801300202887

Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review*, *99*(4), 633–649. https://doi.org/10.1037//0033-295x.99.4.633

Cunningham, C. A., & Egeth, H. E. (2016). Taming the White Bear: Initial Costs and Eventual Benefits of Distractor Inhibition. *Psychological Science*, *27*(4), 476–485. https://doi.org/10.1177/0956797615626564 de Vries, I. E. J., van Driel, J., & Olivers, C. N. L. (2017). Posterior α EEG Dynamics Dissociate Current from Future Goals in Working Memory-Guided Visual Search. *Journal of Neuroscience*, *37*(6), 1591–1603. https://doi.org/10.1523/JNEUROSCI.2945-16.2016

de Waard, J., Bogaerts, L., van Moorselaar, D., & Theeuwes, J. (2022). Surprisingly inflexible: Statistically learned suppression of distractors generalizes across contexts. *Attention, Perception, and Psychophysics*, *84*(2), 459–473. https://doi.org/10.3758/s13414-021-02387-x

Desimone, R. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(1), 193–222. https://doi.org/10.1146/annurev.neuro.18.1.193

Desimone, R. (1998). Visual attention mediated by biased competition in extrastriate visual cortex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *353*(1373), 1245–1255. https://doi.org/10.1098/rstb.1998.0280

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*, 193–222. https://doi.org/10.1146/annurev-psych-122414-033400

DiCarlo, J. J., & Maunsell, J. H. R. (2003). Anterior inferotemporal neurons of monkeys engaged in object recognition can be highly sensitive to object retinal position. *Journal of Neurophysiology*, *89*(6), 3264–3278. https://doi.org/10.1152/jn.00358.2002

Dijkstra, N., Ambrogioni, L., Vidaurre, D., & van Gerven, M. (2020). Neural dynamics of perceptual inference and its reversal during imagery. *eLife*, *9*, 1–19. https://doi.org/10.7554/eLife.53588

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2017). Vividness of visual imagery depends on the neural overlap with perception in visual areas. *Journal of Neuroscience*, *37*(5), 1367–1373. https://doi.org/10.1523/JNEUROSCI.3022-16.2016

Dijkstra, N., Bosch, S. E., & van Gerven, M. A. J. (2019). Shared Neural Mechanisms of Visual Perception and Imagery. *Trends in Cognitive Sciences*, *23*(5), 423–434. https://doi.org/10.1016/j.tics.2019.02.004

Dijkstra, N., Mazor, M., Kok, P., & Fleming, S. (2021). Mistaking imagination for reality: Congruent mental imagery leads to more liberal perceptual detection. *Cognition*, *212*(April), 104719. https://doi.org/10.1016/j.cognition.2021.104719

Dodwell, G., & Eimer, M. (2024). The Preparatory Activation of Guidance Templates for Visual Search and of Target Templates in Non-Search Tasks. *Journal of Cognition*, *7*(1), 1–19. https://doi.org/10.5334/joc.341

Downing, P. E. (2000). Interactions between visual working memory and selective attention. *Psychological Science*, *11*(6). https://doi.org/10.1111/1467-9280.00

Downing, P. E., & Dodds, C. M. (2004). Competition in visual working memory for control of search. *Visual Cognition*, *11*(6), 689–703. https://doi.org/10.1080/13506280344000446

Draschkow, D., & Võ, M. L. H. (2017). Scene grammar shapes the way we interact with objects, strengthens memories, and speeds search. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-16739-x

Duncan, D. H., Moorselaar, D. van, & Theeuwes, J. (2023). Pinging the brain to reveal the hidden attentional priority map using encephalography. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-40405-8

Duncan, J., & Humphreys, G. W. (1989). Visual Search and Stimulus Similarity. *Psychological Review*, *96*(3), 433–458. https://doi.org/10.3758/BF03211797

Eckstein, M. P., Drescher, B. A., & Shimozaki, S. S. (2006). Attentional cues in real scenes, saccadic targeting, and Bayesian priors. *Psychological Science*, *17*(11), 973–980. https://doi.org/10.1111/j.1467-9280.2006.01815.x

Eckstein, M. P., Koehler, K., Welbourne, L. E., & Akbas, E. (2017). Humans, but Not Deep Neural Networks, Often Miss Giant Targets in Scenes. *Current Biology*, *27*(18), 2827-2832.e3. https://doi.org/10.1016/J.CUB.2017.07.068

Ehinger, K. A., Hidalgo-Sotelo, B., Torralba, A., & Oliva, A. (2009). Modelling search for people in 900 scenes: A combined source model of eye guidance. *Visual Cognition*, *17*(6–7), 945–978. https://doi.org/10.1080/13506280902834720

Eimer, M. (2014). The neural basis of attentional control in visual search. *Trends in Cognitive Sciences*, *18*(10), 526–535. https://doi.org/10.1016/j.tics.2014.05.005

Erez, Y., & Yovel, G. (2017). Clutter Modulates the Representation of Target Objects in the Human Occipitotemporal Cortex. *Journal of Cognitive Neuroscience*, *26*(3), 490–500. https://doi.org/10.1162/jocn

Esterman, M., & Yantis, S. (2010). Perceptual expectation evokes category-selective cortical activity. *Cerebral Cortex*, *20*(5), 1245–1253. https://doi.org/10.1093/cercor/bhp188

Fang, M. W. H., & Liu, T. (2019). The profile of attentional modulation to visual features. *Journal of Vision*, *19*(13), 13. https://doi.org/10.1167/19.13.13

Farah, M. J. (1985). Psychophysical Evidence for a Shared Representational Medium for Mental Images and Percepts. *Journal of Experimental Psychology: General*, *114*(1), 91–103. https://doi.org/10.1037/0096-3445.114.1.91

Farah, M. J. (1989). Mechanisms of Imagery-Perception Interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(2), 203–211. https://doi.org/10.1037/0096-1523.15.2.203

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Favila, S. E., & Aly, M. (2023). Hippocampal differentiation of competing memories predicts the precision of preparatory coding in visual cortex. *bioRxiv*. https://doi.org/10.1101/2023.10.09.561548

Fei-Fei, L., Iyer, A., Koch, C., & Perona, P. (2007). What do we perceive in a glance of a real-world scene? *Journal of Vision*, *7*(1), 10. https://doi.org/10.1167/7.1.10

Ferrante, O., Zhigalov, A., Hickey, C., & Jensen, O. (2023). Statistical Learning of Distractor Suppression Downregulates Prestimulus Neural Excitability in Early Visual Cortex. *Journal of Neuroscience*, *43*(12), 2190–2198. https://doi.org/10.1523/JNEUROSCI.1703-22.2022

Finke, R. A. (1986). Some Consequences of Visualization in Pattern Identification and Detection. *The American Journal of Psychology*, *99*(2), 257–274.

Flesch, T., Balaguer, J., Dekker, R., Nili, H., & Summerfield, C. (2018). Comparing continual task learning in minds and machines. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(44), E10313–E10322. https://doi.org/10.1073/pnas.1800755115

Foley, N. C., Kelly, S. P., Mhatre, H., Lopes, M., & Gottlieb, J. (2017). Parietal neurons encode expected gains in instrumental information. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(16), E3315–E3323. https://doi.org/10.1073/pnas.1613844114

Found, A., & Müller, H. J. (1996). Searching for unknown feature targets on more than one dimension: Investigating a "dimension-weighting" account. *Perception and Psychophysics*, *58*(1), 88–101. https://doi.org/10.3758/BF03205479

Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*(3), 134–141. https://doi.org/10.1016/j.tics.2013.01.010

Gandolfo, M., & Downing, P. E. (2019). Causal Evidence for Expression of Perceptual Expectations in Category-Selective Extrastriate Regions. *Current Biology*, *29*(15), 2496-2500.e3. https://doi.org/10.1016/j.cub.2019.06.024

Gaspelin, N., & Luck, S. J. (2018). The Role of Inhibition in Avoiding Distraction by Salient Stimuli. *Trends in Cognitive Sciences*, *22*(1), 79–92. https://doi.org/10.1016/j.tics.2017.11.001

Gayet, S., Battistoni, E., Thorat, S., & Peelen, M. (2024). Searching near and far: The attentional template incorporates viewing distance. *JEP:HPP*, *50*(2), 216–231. https://doi.org/10.1167/jov.23.9.4686

Gayet, S., Guggenmos, M., Christophel, T. B., Haynes, J. D., Paffen, C. L. E., Stigchel, S. V. D., & Sterzer, P. (2017). Visual working memory enhances the neural response to matching visual input. *Journal of Neuroscience*, *37*(28), 6638–6647. https://doi.org/10.1523/JNEUROSCI.3418-16.2017

Gayet, S., Paffen, C. L. E., & Stigchel, S. V. der. (2013). Information Matching the Content of Visual Working Memory Is Prioritized for Conscious Access. *Psychological Science*, *24*(12), 2472–2480. https://doi.org/10.1177/0956797613495882

Gayet, S., Paffen, C. L. E., & Stigchel, S. V. der. (2018). Visual Working Memory Storage Recruits Sensory Processing Areas. *Trends in Cognitive Sciences*, *22*(3), 189–190. https://doi.org/10.1016/j.tics.2017.09.011

Gayet, S., & Peelen, M. V. (2022). Preparatory attention incorporates contextual expectations. *Current Biology*, *32*(3), 687-692.e6. https://doi.org/10.1016/j.cub.2021.11.062

Gazzaley, A., & Nobre, A. C. (2012). Top-down modulation: Bridging selective attention and working memory. *Trends in Cognitive Sciences*, *16*(2), 129–135. https://doi.org/10.1016/j.tics.2011.11.014

Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics*, *67*(7), 1252–1268. https://doi.org/10.3758/BF03193557

Geng, J. J., Di Quattro, N. E., & Helm, J. (2017). Distractor probability changes the shape of the attentional template. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1993–2007. https://doi.org/10.1037/xhp0000430

Geng, J. J., Quattro, N. E. D., & Helm, J. (2017). Distractor probability changes the shape of the attentional template. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(12), 1993–2007. https://doi.org/10.1037/xhp0000430

Geng, J. J., & Witkowski, P. (2019). Template-to-distractor distinctiveness regulates visual search efficiency. *Current Opinion in Psychology*, *29*, 119–125. https://doi.org/10.1016/j.copsyc.2019.01.003

Giesbrecht, B., Weissman, D. H., Woldorff, M. G., & Mangun, G. R. (2006). Pre-target activity in visual cortex predicts behavioral performance on spatial and feature attention tasks. *Brain Research*, *1080*(1), 63–72. https://doi.org/10.1016/j.brainres.2005.09.068

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Golan, A., & Lamy, D. (2024). Attentional guidance by target-location probability cueing is largely inflexible, long-lasting, and distinct from inter-trial priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *50*(2), 244–265. https://doi.org/10.1037/xlm0001220

Gong, M., Chen, Y., & Liu, T. (2022). Preparatory attention to visual features primarily relies on non-sensory representation. *Scientific Reports*, *12*(1), 1–12. https://doi.org/10.1038/s41598-022-26104-2

Gottlieb, J., & Oudeyer, P. Y. (2018). Towards a neuroscience of active sampling and curiosity. *Nature Reviews Neuroscience*, *19*(12), 758–770. https://doi.org/10.1038/s41583-018-0078-0

Gottlieb, J., Oudeyer, P.-Y., Lopes, M., & Baranes, A. (2013). Information seeking, curiosity and attention: Computational and neural mechanisms Information seeking in machine learning, psychology and neuroscience. *Trends in Cognitive Sciences*, *17*(11), 585–593. https://doi.org/10.1016/j.tics.2013.09.001

Greene, Michelle, M., & Oliva, Aude, M. (2009). The Briefest of Glances: The Time Course of Natural Scene Understanding. *Psychological Science*, *20*(4), 464–472.

Grefkes, C., & Fink, G. R. (2005). The functional organization of the intraparietal sulcus in humans and monkeys. *Journal of Anatomy*, *207*(1), 3–17. https://doi.org/10.1111/j.1469-7580.2005.00426.x

Gregoriou, G. G., Gotts, S. J., & Desimone, R. (2012). Cell-Type-Specific Synchronization of Neural Activity in FEF with V4 during Attention. *Neuron*, *73*(3), 581–594. https://doi.org/10.1016/j.neuron.2011.12.019

Gulyás, E., Gombos, F., Sütöri, S., Lovas, A., Ziman, G., & Kovács, I. (2022). Visual imagery vividness declines across the lifespan. *Cortex*, *154*, 365–374. Scopus. https://doi.org/10.1016/j.cortex.2022.06.011

Günseli, E., & Aly, M. (2020). Preparation for upcoming attentional states in the hippocampus and medial prefrontal cortex. *eLife*, *9*, 1–33. https://doi.org/10.7554/elife.53191

Gunseli, E., Meeter, M., & Olivers, C. N. L. (2014). Is a search template an ordinary working memory? Comparing electrophysiological markers of working memory maintenance for visual search and recognition. *Neuropsychologia*, *60*(1), 29–38. https://doi.org/10.1016/j.neuropsychologia.2014.05.012

Hallenbeck, G. E., Sprague, T. C., Rahmati, M., Sreenivasan, K. K., & Curtis, C. E. (2021). Working memory representations in visual cortex mediate distraction effects. *Nature Communications*, *12*(1), 1–18. https://doi.org/10.1038/s41467-021-24973-1

Hamblin-Frohman, Z., & Becker, S. I. (2021). The attentional template in high and low similarity search: Optimal tuning or tuning to relations? *Cognition*, *212*(April), 104732. https://doi.org/10.1016/j.cognition.2021.104732

Hansmann-Roth, S., Kristjánsson, Á., Whitney, D., & Chetverikov, A. (2021). Dissociating implicit and explicit ensemble representations reveals the limits of visual perception and the richness of behavior. *Scientific Reports*, *11*(1), 1–12. https://doi.org/10.1038/s41598-021-83358-y

Harel, A., Kravitz, D. J., & Baker, C. I. (2014). Task context impacts visual object processing differentially across the cortex. *Proceedings of the National Academy of Sciences*, *111*(10), E962–E971. https://doi.org/10.1073/pnas.1312567111

Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635. https://doi.org/10.1038/nature07832

Hebart, M. N., Dickter, A. H., Kidder, A., Kwok, W. Y., Corriveau, A., Van Wicklin, C., & Baker, C. I. (2019). THINGS: A database of 1,854 object concepts and more than 26,000 naturalistic object images. *PLoS ONE*, *14*(10), 1–24. https://doi.org/10.1371/journal.pone.0223792

Hebart, M. N., Görgen, K., & Haynes, J. D. (2015). The decoding toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*(JAN), 1–18. https://doi.org/10.3389/fninf.2014.00088

Helbing, J., Draschkow, D., & L.-H. Võ, M. (2022). Auxiliary Scene-Context Information Provided by Anchor Objects Guides Attention and Locomotion in Natural Search Behavior. *Psychological Science*, *33*(9), 1463–1476. https://doi.org/10.1177/09567976221091838

Henderson, M. M., Serences, J. T., & Rungratsameetaweemana, N. (2023). Categorization dynamically alters representations in human visual cortex. *bioRxiv*, 2023.09.11.557257.

Henriksson, L., Mur, M., & Kriegeskorte, N. (2019). Rapid Invariant Encoding of Scene Layout in Human OPA. *Neuron*, *103*(1), 161-171.e3. https://doi.org/10.1016/j.neuron.2019.04.014

Hetenyi, D., Haarsma, J., & Kok, P. (2024). Pre-stimulus alpha oscillations encode stimulus-specific visual predictions. *bioRxiv*, 2024–03. https://doi.org/10.1101/2024.03.13.584593

Hollingworth, A. (2012). Guidance of visual search by memory and knowledge. In *The influence of attention, learning, and motivation on visual search* (pp. 63–89). Springer. https://doi.org/10.1007/978-1-4614-4794-8_4

Hong, H., Yamins, D. L. K., Majaj, N. J., & Dicarlo, J. J. (2016). Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, *19*(4), 613–622. https://doi.org/10.1038/nn.4247

Hout, M. C., & Goldinger, S. D. (2014). Target templates: The precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, and Psychophysics*, *77*(1), 128–149. https://doi.org/10.3758/s13414-014-0764-6

Hu, Y., & Yu, Q. (2023). Spatiotemporal dynamics of self-generated imagery reveal a reverse cortical hierarchy from cue-induced imagery. *Cell Reports*, *42*(10), 2023.01.25.525474. https://doi.org/10.1016/j.celrep.2023.113242

Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences*, *16*(12), 576–579. https://doi.org/10.1016/j.tics.2012.10.003

Iamshchinina, P., Christophel, T. B., Gayet, S., & Rademaker, R. L. (2021). Essential considerations for exploring visual working memory storage in the human brain. *Visual Cognition*, *29*(7), 1–12. https://doi.org/10.1080/13506285.2021.1915902

Iamshchinina, P., Kaiser, D., Yakupov, R., Haenelt, D., Sciarra, A., Mattern, H., Luesebrink, F., Duezel, E., Speck, O., Weiskopf, N., & Cichy, R. M. (2021). Perceived and mentally rotated contents are differentially represented in cortical depth of V1. *Communications Biology*, *4*(1), 1–8. https://doi.org/10.1038/s42003-021-02582-4

Ishai, A., & Sagi, D. (1987). Visual Imagery: Effects of Short- and Long-Term Memory. *Journal of Cognitive Neuroscience*, *9*(6), 734–742.

Ishai, A., & Sagi, D. (1995). Common Mechanisms of Visual Imagery and Perception. *Science*, *268*(5218), 1772–1774. https://doi.org/10.1126/science.7792605

Ishai, A., Ungerleider, L. G., & Haxby, J. V. (2000). Distributed neural systems for the generation of visual images. *Neuron*, *28*(3), 979–990. https://doi.org/10.1016/S0896-6273(00)00168-9

Jazayeri, M., & Movshon, J. A. (2006). Optimal representation of sensory information by neural populations. *Nature Neuroscience*, *9*(5), 690–696. https://doi.org/10.1038/nn1691

Julian, J. B., Fedorenko, E., Webster, J., & Kanwisher, N. (2012). An algorithmic method for functionally defining regions of interest in the ventral visual pathway. *NeuroImage*, *60*(4), 2357–2364. https://doi.org/10.1016/j.neuroimage.2012.02.055

Kaiser, D., & Cichy, R. M. (2018). Typical visual-field locations facilitate access to awareness for everyday objects. *Cognition*, *180*(July), 118–122. https://doi.org/10.1016/j.cognition.2018.07.009

Kaiser, D., Oosterhof, N. N., & Peelen, M. V. (2016). The Neural Dynamics of Attentional Selection in Natural Scenes. *Journal of Neuroscience*, *36*(41), 10522–10528. https://doi.org/10.1523/JNEUROSCI.1385-16.2016

Kaiser, D., & Peelen, M. V. (2018). Transformation from independent to integrative coding of multi-object arrangements in human visual cortex. *NeuroImage*, *169*(March 2017), 334–341. https://doi.org/10.1016/j.neuroimage.2017.12.065

Kaiser, D., Quek, G. L., Cichy, R. M., & Peelen, M. V. (2019). Object Vision in a Structured World. *Trends in Cognitive Sciences*, *23*(8), 672–685. https://doi.org/10.1016/j.tics.2019.04.013

Kaiser, D., Stein, T., & Peelen, M. V. (2014). Object grouping based on real-world regularities facilitates perception by reducing competitive interactions in visual cortex. *Proceedings of the National Academy of Sciences*, *111*(30), 11217–11222. https://doi.org/10.1073/pnas.1400559111

Kallmayer, A., Zacharias, L., Jetter, L., & Võ, M. L. (2024). *More than meets the eye: Neural evidence for scene grammar representations during individual object processing* . PsyArXiv. https://osf.io/preprints/psyarxiv/hs835

Kastner, S., De Weerd, P., Pinsk, M. A., Elizondo, M. I., Desimone, R., & Ungerleider, L. G. (2001). Modulation of sensory suppression: Implications for receptive field sizes in the human visual cortex. *Journal of Neurophysiology*, *86*(3), 1398–1411. https://doi.org/10.1152/jn.2001.86.3.1398

Kastner, S., Pinsk, M. A., Weerd, P. D., Desimone, R., & Ungerleider, L. G. (1999). Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation. *Neuron*, *22*(4), 751–761. https://doi.org/10.1016/S0896-6273(00)80734-5

Keogh, R., & Pearson, J. (2011). Mental imagery and visual working memory. *PLoS ONE*, *6*(12). https://doi.org/10.1371/journal.pone.0029221

Keogh, R., & Pearson, J. (2014). The sensory strength of voluntary visual imagery predicts visual working memory capacity. *Journal of Vision*, *14*(12), 1–13. https://doi.org/10.1167/14.12.7

Kerkoerle, T. V., Self, M. W., & Roelfsema, P. R. (2017). Layer-specificity in the effects of attention and working memory on activity in primary visual cortex. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms13804

Kershner, A. M., & Hollingworth, A. (2022). Real-world object categories and scene contexts conjointly structure statistical learning for the guidance of visual search. *Attention, Perception, and Psychophysics*, *84*(4), 1304–1316. https://doi.org/10.3758/s13414-022-02475-6

Kerzel, D. (2020). Direct Evidence for the Optimal Tuning of Attention. *Journal of Experimental Psychology: Human Perception and Performance*, *46*(7), 716–728. https://doi.org/10.1037/xhp0000744

Kiyonaga, A., Dowd, E. W., & Egner, T. (2017). Neural Representation of Working Memory Content Is Modulated by Visual Attentional Demand. *Journal of Cognitive Neuroscience*, *29*(12), 2011–2024. https://doi.org/10.1162/jocn_a_01174

Kiyonaga, A., & Egner, T. (2016). Center-Surround Inhibition in Working Memory. *Current Biology*, *26*(1), 64–68. https://doi.org/10.1016/j.cub.2015.11.013

Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147. https://doi.org/10.1016/S1364-6613(00)01452-2

Kliger, L., & Yovel, G. (2020). The functional organization of high-level visual cortex determines the representation of complex visual stimuli. *Journal of Neuroscience*, *40*(39), 7545–7558. https://doi.org/10.1523/JNEUROSCI.0446-20.2020

Koehler, K., & Eckstein, M. P. (2017a). Beyond scene gist: Objects guide search more than scene background. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(6), 1177–1193. https://doi.org/10.1037/xhp0000363

Koehler, K., & Eckstein, M. P. (2017b). Temporal and peripheral extraction of contextual cues from scenes during visual search. *Journal of Vision*, *17*(2). https://doi.org/10.1167/17.2.6

Koenig-Robert, R., & Pearson, J. (2021). Why do imagery and perception look and feel so different ? *Philosophical Transactions of the Royal Society B: Biological Sciences*, *376*(1817), 20190703.

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior Expectations Evoke Stimulus Templates in the Primary Visual Cortex. *Journal of Cognitive Neuroscience*, *26*(7), 1546–1554. https://doi.org/10.1162/jocn_a_00562

Kok, P., Mostert, P., & de Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(39), 10473–10478. https://doi.org/10.1073/pnas.1705652114

Konen, C. S., & Kastner, S. (2008). Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, *11*(2), 224–231. https://doi.org/10.1038/nn2036

Konkle, T., & Caramazza, A. (2013). Tripartite organization of the ventral stream by animacy and object size. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(25), 10235–10242. https://doi.org/10.1523/JNEUROSCI.0983-13.2013

Kosslyn, S. M., Ganis, G., & Thompson, W. L. (2001). Neural foundations of imagery. *Nature Reviews Neuroscience*, *2*(9), 635–642. https://doi.org/10.1038/35090055

Kourtzi, Z., & Connor, C. E. (2011). Neural Representations for Object Perception: Structure, Category, and Adaptive Coding. *Annual Review of Neuroscience*, *34*(1), 45–67. https://doi.org/10.1146/annurev-neuro-060909-153218

Kriegeskorte, N. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*. https://doi.org/10.3389/neuro.06.004.2008

Kristjánsson, Á. (2022). Priming of probabilistic attentional templates. *Psychonomic Bulletin & Review*, *0123456789*, 1–36. https://doi.org/10.3758/s13423-022-02125-w

Kumar, S., Soto, D., & Humphreys, G. W. (2009). Electrophysiological evidence for attentional guidance by the contents of working memory. *European Journal of Neuroscience*, *30*(2), 307–317. https://doi.org/10.1111/j.1460-9568.2009.06805.x

Lau, J. S. H., Pashler, H., & Brady, T. F. (2021). Target templates in low target-distractor discriminability visual search have higher resolution, but the advantage they provide is short-lived. *Attention, Perception, and Psychophysics*. https://doi.org/10.3758/s13414-020-02213-w

Lee, J., & Geng, J. J. (2020). Flexible weighting of target features based on distractor context. *Attention, Perception, and Psychophysics*, *82*(2), 739–751. https://doi.org/10.3758/s13414-019-01910-5

Lee, S. H., Kravitz, D. J., & Baker, C. I. (2012). Disentangling visual imagery and perception of real-world objects. *NeuroImage*, *59*(4), 4064–4073. https://doi.org/10.1016/j.neuroimage.2011.10.055

Li, F. F., VanRullen, R., Koch, C., & Perona, P. (2002). Rapid natural scene categorization in the near absence of attention. *Proceedings of the National Academy of Sciences*, *99*(14), 9596–9601. https://doi.org/10.1073/pnas.092277599

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for Participants' Viewing Distance in Large-Scale, Psychophysical Online Experiments Using a Virtual Chinrest. *Scientific Reports*, *10*(1), 1–11. https://doi.org/10.1038/s41598-019-57204-1

Liesefeld, H. R., Liesefeld, A. M., Pollmann, S., & Müller, H. J. (2019). Biasing Allocations of Attention via Selective Weighting of Saliency Signals: Behavioral and Neuroimaging Evidence for the Dimension-Weighting Account. In *Current Topics in Behavioral Neurosciences* (Vol. 41, pp. 87–113). Springer. https://doi.org/10.1007/7854_2018_75

Liu, H., Agam, Y., Madsen, J. R., & Kreiman, G. (2009). Timing, Timing, Timing: Fast Decoding of Object Information from Intracranial Field Potentials in Human Visual Cortex. *Neuron*, *62*(2), 281–290. https://doi.org/10.1016/j.neuron.2009.02.025

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-Based Attention Modulates Orientation-Selective Responses in Human Visual Cortex. *Neuron*, *55*(2), 313–323. https://doi.org/10.1016/j.neuron.2007.06.030

Lleras, A., Buetti, S., & Xu, Z. J. (2022). Incorporating the properties of peripheral vision into theories of visual search. *Nature Reviews Psychology*, *1*(October), 590–604. https://doi.org/10.1038/s44159-022-00097-1

Loon, A. M. van, Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. L. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *eLife*, *7*, 1–25. https://doi.org/10.7554/eLife.38677

Lorenc, E. S., Mallett, R., & Lewis-Peacock, J. A. (2021). Distraction in Visual Working Memory: Resistance is Not Futile. *Trends in Cognitive Sciences*, *25*(3), 228–239. https://doi.org/10.1016/j.tics.2020.12.004

Lorenc, E. S., Sreenivasan, K. K., Nee, D. E., Vandenbroucke, A. R. E., & D'Esposito, M. (2018). Flexible coding of visual working memory representations during distraction. *Journal of Neuroscience*, *38*(23), 5267–5276. https://doi.org/10.1523/JNEUROSCI.3061-17.2018

Luck, S. J. (2012). Electrophysiological Correlates of the Focusing of Attention within Complex Visual Scenes: N2pc and Related ERP Components. In *The Oxford Handbook of Event-Related Potential Components*. https://doi.org/10.1093/oxfordhb/9780195374148.013.0161

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. https://doi.org/10.1152/jn.1997.77.1.24

Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, *4*(11), 432–440. https://doi.org/10.1016/S1364-6613(00)01545-X

MacEvoy, S. P., & Epstein, R. A. (2009). Decoding the Representation of Multiple Simultaneous Objects in Human Occipitotemporal Cortex. *Current Biology*, *19*(11), 943–947. https://doi.org/10.1016/j.cub.2009.04.020

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, *11*(9), 1–16. https://doi.org/10.1167/11.9.9

Madden, D. J., Whiting, W. L., Cabeza, R., & Huettel, S. A. (2004). Age-Related Preservation of Top-Down Attentional Guidance During Visual Search. *Psychology and Aging*, *19*(2), 304–309. https://doi.org/10.1037/0882-7974.19.2.304

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision*, *9*(11), 1–13. https://doi.org/10.1167/9.11.8

Maljkovic, V., & Nakayama, K. (1994). Priming of pop-out: 1. Role of features. *Memory & Cognition*, *22*(6), 657–672.

Marks, D. F. (1973). Visual Imagery Differences in the Recall of Pictures. *British Journal of Psychology*, *64*(1), 17–24. https://doi.org/10.1111/j.2044-8295.1973.tb01322.x

Martinez-Trujillo, J. (2022). Visual Attention in the Prefrontal Cortex. *Annual Review of Vision Science*, *8*(1), 1–19. https://doi.org/10.1146/annurev-vision-100720-031711

Maunsell, J. H. R., & Treue, S. (2006). Feature-based attention in visual cortex. *Trends in Neurosciences*, *29*(6), 317–322. https://doi.org/10.1016/j.tins.2006.04.001

Maxfield, J. T., & Zelinsky, G. J. (2012). Searching through the hierarchy: How level of target categorization affects visual search. *Visual Cognition*, *20*(10), 1153–1163. https://doi.org/10.1080/13506285.2012.735718

McKee, J. L., Riesenhuber, M., Miller, E. K., & Freedman, D. J. (2014). Task dependence of visual and category representations in prefrontal and inferior temporal cortices. *Journal of Neuroscience*, *34*(48), 16065–16075. https://doi.org/10.1523/JNEUROSCI.1660-14.2014

Mendoza-Halliday, D., Xu, H., Azevedo, F. A. C., & Desimone, R. (2024). Dissociable neuronal substrates of visual feature attention and working memory. *Neuron*, *112*(5), 850-863.e6. https://doi.org/10.1016/j.neuron.2023.12.007

Miller, E. K., Erickson, C. A., & Desimone, R. (1996). Neural Mechanisms of Visual Working Memory in Prefrontal Cortex of the Macaque. *Journal of Neuroscience*, *16*(16), 5154–5167. https://doi.org/10.1523/JNEUROSCI.16-16-05154.1996

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

Moerel, D., Grootswagers, T., Robinson, A. K., Shatek, S. M., Woolgar, A., Carlson, T. A., & Rich, A. N. (2022). The time-course of feature-based attention effects dissociated from temporal expectation and target-related processes. *Scientific Reports*, *12*(1), 1–14. https://doi.org/10.1038/s41598-022-10687-x

Moher, J., & Egeth, H. E. (2012). The ignoring paradox: Cueing distractor features leads first to selection, then to inhibition of to-be-ignored items. *Attention, Perception, & Psychophysics*, *74*(8), 1590–1605. https://doi.org/10.3758/s13414-012-0358-0

Mongillo, G., Barak, O., & Tsodyks, M. (2008). Synaptic Theory of Working Memory. *Science*, *319*(5869), 1543–1546. https://doi.org/10.1126/science.1150769

Moore, T., & Armstrong, K. M. (2003). Selective gating of visual signals by microstimulation of frontal cortex. *Nature*, *421*(6921), 370–373. https://doi.org/10.1038/nature01341

Moorselaar, D. van, Gayet, S., Paffen, C. L. E., Theeuwes, J., Stigchel, S. V. der, & Olivers, C. N. L. (2018). Competitive interactions in visual working memory drive access to awareness. *Cortex*, *102*, 6–13. https://doi.org/10.1016/J.CORTEX.2017.03.026

Moriya, J. (2018). Visual mental imagery influences attentional guidance in a visual-search task. *Attention, Perception, & Psychophysics*, *80*(5), 1127–1142. https://doi.org/10.3758/s13414-018-1520-0

Myers, N. E., Rohenkohl, G., Wyart, V., Woolrich, M. W., Nobre, A. C., & Stokes, M. G. (2015). Testing sensory evidence against mnemonic templates. *eLife*, *4*(DECEMBER2015), 1–25. https://doi.org/10.7554/eLife.09000

Naselaris, T., Olman, C. A., Stansbury, D. E., Ugurbil, K., & Gallant, J. L. (2015). A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *NeuroImage*, *105*, 215–228. https://doi.org/10.1016/j.neuroimage.2014.10.018

Navalpakkam, V., & Itti, L. (2007). Search Goal Tunes Visual Features Optimally. *Neuron*, *53*(4), 605–617. https://doi.org/10.1016/j.neuron.2007.01.018

Neider, M. B., & Zelinsky, G. J. (2006). Scene context guides eye movements during visual search. *Vision Research*, *46*(5), 614–621. https://doi.org/10.1016/j.visres.2005.08.025

Niklaus, M., Nobre, A. C., & Van Ede, F. (2017). Feature-based attentional weighting and spreading in visual working memory. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/srep42384

Noah, S., Powell, T., Khodayari, N., Olivan, D., Ding, M., & Mangun, G. R. (2020). Neural Mechanisms of Attentional Control for Objects: Decoding EEG Alpha When Anticipating Faces, Scenes,and Tools. *Journal of Neuroscience*, *40*(25), 4913–4924. https://doi.org/10.1523/JNEUROSCI.2685-19.2020

Nobre, A. C., & Serences, J. T. (2018). Building on a Solid Baseline: Anticipatory Biases in Attention. *Trends in Neurosciences*, *41*(3), 117–120. https://doi.org/10.1016/j.tins.2017.12.005

O'Craven, K. M., & Kanwisher, N. (2000). Mental imagery of faces and places activates corresponding stimulus-specific brain regions. *Journal of Cognitive Neuroscience*, *12*(6), 1013–1023. https://doi.org/10.1162/08989290051137549

Olivers, C. N. L. (2009). What Drives Memory-Driven Attentional Capture? The Effects of Memory Type, Display Type, and Search Type. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(5), 1275–1291. https://doi.org/10.1037/a0013896

Olivers, C. N. L., & Eimer, M. (2011). On the difference between working memory and attentional set. *Neuropsychologia*, *49*(6), 1553–1558. https://doi.org/10.1016/j.neuropsychologia.2010.11.033

Olivers, C. N. L., Meijer, F., & Theeuwes, J. (2006). Feature-based memory-driven attentional capture: Visual working memory content affects visual attention. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1243.

Olivers, C. N. L., Peters, J., Houtkamp, R., & Roelfsema, P. R. (2011). Different states in visual working memory: When it guides attention and when it does not. *Trends in Cognitive Sciences*, *15*(7), 327–334. https://doi.org/10.1016/j.tics.2011.05.004

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, *2011*. https://doi.org/10.1155/2011/156869

Oosterhof, N. N., Connolly, A. C., & Haxby, J. V. (2016). CoSMoMVPA: Multi-modal multivariate pattern analysis of neuroimaging data in matlab/GNU octave. *Frontiers in Neuroinformatics*, *10*(JUL). https://doi.org/10.3389/fninf.2016.00027

Park, Y. E., Sy, J. L., Hong, S. W., & Tong, F. (2017). Reprioritization of Features of Multidimensional Objects Stored in Visual Working Memory. *Psychological Science*, *28*(12), 1773–1785. https://doi.org/10.1177/0956797617719949

Pashler, H., & Shiu, L. P. (1999). Do images involuntarily trigger search? A test of Pillsbury's hypothesis. *Psychonomic Bulletin and Review*, *6*(3), 445–448. https://doi.org/10.3758/BF03210833

Pearson, J. (2019). The human imagination: The cognitive neuroscience of visual mental imagery. *Nature Reviews Neuroscience*, *20*(10), 624–634. https://doi.org/10.1038/s41583-019-0202-9

Pearson, J., Clifford, C. W. G., & Tong, F. (2008). The functional impact of mental imagery on conscious perception. *Current Biology : CB*, *18*(13), 982–986. https://doi.org/10.1016/j.cub.2008.05.048

Peelen, M. V., Berlot, E., & de Lange, F. P. (2023). Predictive processing of scenes and objects. *Nature Reviews Psychology*, *725970*, 1–27. https://doi.org/10.1038/s44159-023-00254-0

Peelen, M. V., He, C., Han, Z., Caramazza, A., & Bi, Y. (2014). Nonvisual and Visual Object Shape Representations in Occipitotemporal Cortex: Evidence from Congenitally Blind and Sighted Adults. *Journal of Neuroscience*, *34*(1), 163–170. https://doi.org/10.1523/JNEUROSCI.1114-13.2014

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences*, *108*(29), 12125–12130. https://doi.org/10.1073/pnas.1101042108

Peelen, M. V., & Kastner, S. (2014). Attention in the real world: Toward understanding its neural basis. *Trends in Cognitive Sciences*, *18*(5), 242–250. https://doi.org/10.1016/J.TICS.2014.02.004

Peelen, M. V., Li, F. F., & Kastner, S. (2009). Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature*, *7251*(420), 94–97. https://doi.org/10.1016/j.physbeh.2017.03.040

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Pereira, E. J., & Castelhano, M. S. (2019). Attentional capture is contingent on scene region: Using surface guidance framework to explore attentional mechanisms during search. *Psychonomic Bulletin & Review*, *26*(4), 1273–1281. https://doi.org/10.3758/s13423-019-01610-z

Potter, M. C. (1975). Meaning in Visual Search. *Science*, *187*(4180), 965–966. https://doi.org/10.1126/science.1145183

Pouget, A., & Bavelier, D. (2007). Paying Attention to Neurons with Discriminating Taste. *Neuron*, *53*(4), 473–475. https://doi.org/10.1016/j.neuron.2007.02.004

Preston, T. J., Guo, F., Das, K., Giesbrecht, B., & Eckstein, M. P. (2013). Neural representations of contextual guidance in visual search of real-world scenes. *Journal of Neuroscience*, *33*(18), 7846–7855. https://doi.org/10.1523/JNEUROSCI.5840-12.2013

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research*, *1301*, 89–99. https://doi.org/10.1016/j.brainres.2009.08.085

Purushothaman, G., & Bradley, D. C. (2005). Neural population code for fine perceptual decisions in area MT. *Nature Neuroscience*, *8*(1), 99–106. https://doi.org/10.1038/nn1373

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*(8), 1336–1344. https://doi.org/10.1038/s41593-019-0428-x

Rademaker, R. L., & Serences, J. (2024). *Manipulating attentional priority creates a trade-off between memory and sensory representations in human visual cortex* (p. 2024.09.16.613302). bioRxiv. https://doi.org/10.1101/2024.09.16.613302

Rademaker, R. L., & Serences, J. T. (2017). Pinging the brain to reveal hidden memories. *Nature Neuroscience*, *20*(6), 767–769. https://doi.org/10.1038/nn.4560

Reddy, L., & Kanwisher, N. (2007). Category Selectivity in the Ventral Visual Pathway Confers Robustness to Clutter and Diverted Attention. *Current Biology*, *17*(23), 2067–2072. https://doi.org/10.1016/j.cub.2007.10.043

Reddy, L., Kanwisher, N. G., & VanRullen, R. (2009). Attention and biased competition in multi-voxel object representations. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(50), 21447–21452. https://doi.org/10.1073/pnas.0907330106

Reeder, R. R., & Peelen, M. V. (2013). The contents of the search template for category-level search in natural scenes. *Journal of Vision*, *13*(3), 13–13. https://doi.org/10.1167/13.3.13

Reeder, R. R., Perini, F., & Peelen, M. V. (2015). Preparatory activity in posterior temporal cortex causally contributes to object detection in scenes. *Journal of Cognitive Neuroscience*, *27*(11), 2117–2125. https://doi.org/10.1162/jocn_a_00845

Reynolds, J. H., Chelazzi, L., & Desimone, R. (1999). Competitive Mechanisms Subserve Attention in Macaque Areas V2 and V4. *The Journal of Neuroscience*, *19*(5), 1736–1753. https://doi.org/10.1523/JNEUROSCI.19-05-01736.1999

Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168–185. https://doi.org/10.1016/j.neuron.2009.01.002

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2003). Taking the MAX from neuronal responses. *Trends in Cognitive Sciences*, *7*(3), 99–102. https://doi.org/10.1016/S1364-6613(03)00023-8

Rousselet, G. A., Thorpe, S. J., & Fabre-Thorpe, M. (2004). How parallel is visual processing in the ventral pathway? *Trends in Cognitive Sciences*, *8*(8), 363–370. https://doi.org/10.1016/j.tics.2004.06.003

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., Rees, G., Josephs, O., Deichmann, R., & Driver, J. (2006). Concurrent TMS-fMRI and Psychophysics Reveal Frontal Influences on Human Retinotopic Visual Cortex. *Current Biology*, *16*(15), 1479–1488. https://doi.org/10.1016/j.cub.2006.06.057

Saenz, M., Buracas, G. T., & Boynton, G. M. (2002). Global effects of feature-based attention in human visual cortex. *Nature Neuroscience*, *5*(7), 631–632. https://doi.org/10.1038/nn876

Sato, T. (1989). Interactions of visual stimuli in the receptive fields of inferior temporal neurons in awake macaques. *Experimental Brain Research*, *77*(1), 23–30. https://doi.org/10.1007/BF00250563

Schapiro, A. C., Kustner, L. V., & Turk-Browne, N. B. (2012). Shaping of object representations in the human medial temporal lobe based on temporal regularities. *Current Biology*, *22*(17), 1622–1627. https://doi.org/10.1016/j.cub.2012.06.056

Schmidt, J., & Zelinsky, G. J. (2017). Adding details to the attentional template offsets search difficulty: Evidence from contralateral delay activity. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(3), 429–437. https://doi.org/10.1037/xhp0000367

Schorr, D., Balzano, G., & Smith, E. E. (1978). Selective interference between imagery and perception: Is it modality specific or relation specific? *Bulletin of the Psychonomic Society*, *12*(6), 419–422. https://doi.org/10.3758/BF03329725

Schulz, M.-C., Bartsch, M. V., Merkel, C., Strumpf, H., Schoenfeld, M. A., & Hopf, J.-M. (2024). Parallel gain modulation mechanisms set the resolution of color selectivity in human visual cortex. *Science Advances*, *10*(37), eadm7385. https://doi.org/10.1126/sciadv.adm7385

Schwarzlose, R. F., Swisher, J. D., Dang, S., & Kanwisher, N. (2008). The distribution of category and location information across object-selective regions in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(11), 4447–4452. https://doi.org/10.1073/pnas.0800431105

Scolari, M., Byers, A., & Serences, J. T. (2012). Optimal deployment of attentional gain during fine discriminations. *Journal of Neuroscience*, *32*(22), 7723–7733. https://doi.org/10.1523/JNEUROSCI.5558-11.2012

Scolari, M., & Serences, J. T. (2009). Adaptive allocation of attentional gain. *Journal of Neuroscience*, *29*(38), 11933–11942. https://doi.org/10.1523/JNEUROSCI.5642-08.2009

Scolari, M., & Serences, J. T. (2010). Basing perceptual decisions on the most informative sensory neurons. *Journal of Neurophysiology*, *104*(4), 2266–2273. https://doi.org/10.1152/jn.00273.2010

Seeliger, K., Fritsche, M., Güçlü, U., Schoenmakers, S., Schoffelen, J. M., Bosch, S. E., & van Gerven, M. A. J. (2018). Convolutional neural network-based encoding and decoding of visual object recognition in space and time. *NeuroImage*, *180*(July 2017), 253–266. https://doi.org/10.1016/j.neuroimage.2017.07.018

Segal, S., & Fusella, V. (1970). Influence of Imaged Pictures. *Journal of Experimental Psychology*, *83*(3), 458–464.

Serences, J. T., & Boynton, G. M. (2007). Feature-Based Attentional Modulations in the Absence of Direct Visual Stimulation. *Neuron*, *55*(2), 301–312. https://doi.org/10.1016/j.neuron.2007.06.015

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. https://doi.org/10.1111/j.1467-9280.2009.02276.x

Sha, L., Haxby, J. V., Abdi, H., Guntupalli, S. J., Oosterhof, N. N., Halchenko, Y. O., & Connolly, A. C. (2015). The Animacy Continuum in the Human Ventral Vision Pathway. *Journal of Cognitive Neuroscience*, *27*(4), 665–678. https://doi.org/10.1162/jocn

Shang, L., Yeh, L.-C., Zhao, Y., Wiegand, I., & Peelen, M. V. (2024). Category-Based Attention Facilitates Memory Search. *Eneuro*, *11*(2), ENEURO.0012-24.2024. https://doi.org/10.1523/ENEURO.0012-24.2024

Shatek, S. M., Grootswagers, T., Robinson, A. K., & Carlson, T. A. (2019). Decoding images in the mind's eye: The temporal dynamics of visual imagery. *Vision (Switzerland)*, *3*(4), 1–16. https://doi.org/10.3390/vision3040053

Sigala, N., & Logothetis, N. K. (2002). Visual categorization shapes feature selectivity in the primate temporal cortex. *Nature*, *415*(6869), 318–320. https://doi.org/10.1038/415318a

Sisk, C. A., Remington, R. W., & Jiang, Y. V. (2019). Mechanisms of contextual cueing: A tutorial review. *Attention, Perception, and Psychophysics*, *81*(8), 2571–2589. https://doi.org/10.3758/s13414-019-01832-2

Smith, S. M., Glenberg, A., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition*, *6*(4), 342–353. https://doi.org/10.3758/BF03197465

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage*, *44*(1), 83–98. https://doi.org/10.1016/j.neuroimage.2008.03.061

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review*, *8*(2), 203–220.

Solomon, S., Kay, K. N., & Schapiro, A. C. (2024). Recent statistics shift object representations in parahippocampal cortex. *bioRxiv*. https://doi.org/10.1101/2024.02.07.579310

Soon, C. S., Namburi, P., & Chee, M. W. L. (2013). Preparatory patterns of neural activity predict visual category search speed. *NeuroImage*, *66*, 215–222. https://doi.org/10.1016/j.neuroimage.2012.10.036

Soto, D., Heinke, D., Humphreys, G. W., & Blanco, M. J. (2005). Early, involuntary top-down guidance of attention from working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(2), 248–261. https://doi.org/10.1037/0096-1523.31.2.248

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. https://doi.org/10.1016/j.tics.2008.05.007

Spaak, E., & de Lange, F. P. (2020). Hippocampal and prefrontal theta-band mechanisms underpin implicit spatial context learning. *Journal of Neuroscience*, *40*(1), 191–202. https://doi.org/10.1523/JNEUROSCI.1660-19.2019

Spotorno, S., Malcolm, G. L., & Tatler, B. W. (2014). How context information and target information guide the eyes from the first epoch of search in real-world scenes. *Journal of Vision*, *14*(2), 1–21. https://doi.org/10.1167/14.2.7

Sreenivasan, K. K., & D'Esposito, M. (2019). The what, where and how of delay activity. *Nature Reviews Neuroscience*, *20*(8), 466–481. https://doi.org/10.1038/s41583-019-0176-7

Stecher, R., & Kaiser, D. (2024). Representations of imaginary scenes and their properties in cortical alpha activity. *Scientific Reports*, *14*(1), 1–12. https://doi.org/10.1038/s41598-024-63320-4

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of Experimental Psychology: General*, *144*(6), 1089–1104. https://doi.org/10.1037/xge0000109

Stilwell, B. T., Bahle, B., & Vecera, S. P. (2019). Feature-based statistical regularities of distractors modulate attentional capture. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 419–433. https://doi.org/10.1037/xhp0000613

Stokes, M. G. (2011). Top-down visual activity underlying VSTM and preparatory attention. *Neuropsychologia*, *49*(6), 1425–1427. https://doi.org/10.1016/j.neuropsychologia.2011.02.004

Stokes, M. G. (2015). "Activity-silent" working memory in prefrontal cortex: A dynamic coding framework. *Trends in Cognitive Sciences*, *19*(7), 394–405. https://doi.org/10.1016/j.tics.2015.05.004

Stokes, M. G., Atherton, K., Patai, E. Z., & Nobre, A. C. (2012). Long-term memory prepares neural activity for perception. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(6). https://doi.org/10.1073/pnas.1108555108

Stokes, M. G., Thompson, R., Cusack, R., & Duncan, J. (2009). Top-down activation of shape-specific population codes in visual cortex during mental imagery. *Journal of Neuroscience*, *29*(5), 1565–1572. https://doi.org/10.1523/JNEUROSCI.4657-08.2009

Stokes, M. G., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19569–19574. https://doi.org/10.1073/pnas.0905306106

Stoletniy, A. S., Alekseeva, D. S., Babenko, V. V., Anokhina, P. V., & Yavna, D. V. (2022). The N2pc Component in Studies of Visual Attention. *Neuroscience and Behavioral Physiology*, *52*(8), 1299–1309. https://doi.org/10.1007/s11055-023-01359-y

Störmer, V. S., & Alvarez, G. A. (2014). Feature-Based Attention Elicits Surround Suppression in Feature Space. *Current Biology*, *24*(17), 1985–1988. https://doi.org/10.1016/j.cub.2014.07.030

Störmer, V. S., Cohen, M. A., & Alvarez, G. A. (2019). Tuning Attention to Object Categories: Spatially Global Effects of Attention to Faces in Visual Processing. *Journal of Cognitive Neuroscience*, *31*(7), 937–947. https://doi.org/10.1162/jocn_a_01400

Sulfaro, A. A., Robinson, A. K., & Carlson, T. A. (2023). Modelling perception as a hierarchical competition differentiates imagined, veridical, and hallucinated percepts. *Neuroscience of Consciousness*, *2023*(1). https://doi.org/10.1093/nc/niad018

Sulfaro, A. A., Robinson, A. K., & Carlson, T. A. (2024). Neural interference between real and imagined visual stimuli. *bioRxiv*. https://doi.org/10.1101/2024.01.05.574285

Summerfield, C., & de Lange, F. P. (2014). Expectation in perceptual decision making: Neural and computational mechanisms. *Nature Reviews Neuroscience*, *15*(11), 745–756. https://doi.org/10.1038/nrn3838

Summerfield, J. J., Lepsien, J., Gitelman, D. R., Mesulam, M. M., & Nobre, A. C. (2006). Orienting attention based on long-term memory experience. *Neuron*, *49*(6), 905–916. https://doi.org/10.1016/j.neuron.2006.01.021

Theeuwes, J., Bogaerts, L., & van Moorselaar, D. (2022). What to expect where and when: How statistical learning drives visual selection. *Trends in Cognitive Sciences*, *26*(10), 860–872. https://doi.org/10.1016/j.tics.2022.06.001

Thorat, S., & Peelen, M. V. (2022). Body shape as a visual feature: Evidence from spatially-global attentional modulation in human visual cortex. *NeuroImage*, *255*(December 2021), 119207. https://doi.org/10.1016/j.neuroimage.2022.119207

Thorat, S., Proklova, D., & Peelen, M. V. (2019). The nature of the animacy organization in human ventral temporal cortex. *eLife*, *8*, 1–18. https://doi.org/10.7554/eLife.47142

Thorat, S., Quek, G. L., & Peelen, M. V. (2022). Statistical learning of distractor co-occurrences facilitates visual search. *Journal of Vision*, *22*(10), 1–11. https://doi.org/10.1167/jov.22.10.2

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature*, *381*.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual guidance of eye movements and attention in real-world scenes: The role of global features in object search. *Psychological Review*, *113*(4), 766–786. https://doi.org/10.1037/0033-295X.113.4.766

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

Treue, S. (2014). Attentional Selection: Mexican Hats Everywhere. *Current Biology*, *24*(18), R838–R839. https://doi.org/10.1016/j.cub.2014.08.019

Treue, S., & Martínez Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575–579. Scopus. https://doi.org/10.1038/21176

Turatto, M., Bonetti, F., Pascucci, D., & Chelazzi, L. (2018). Desensitizing the attention system to distraction while idling: A new latent learning phenomenon in the visual attention domain. *Journal of Experimental Psychology: General*, *147*(12), 1827–1850. https://doi.org/10.1037/xge0000503

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, *15*(1), 273–289. https://doi.org/10.1006/nimg.2001.0978

van der Stigchel, S., & Hollingworth, A. (2018). Visuospatial Working Memory as a Fundamental Component of the Eye Movement System. *Current Directions in Psychological Science*, *27*(2), 136–143. https://doi.org/10.1177/0963721417741710

van Ede, F., & Nobre, A. C. (2023). Turning Attention Inside Out: How Working Memory Serves Behavior. *Annual Review of Psychology*, *74*, 137–165. https://doi.org/10.1146/annurev-psych-021422041757 van Lieshout, L. L., de Lange, F. P., & Cools, R. (2020). Why so curious? Quantifying mechanisms of information seeking. *Current Opinion in Behavioral Sciences*, *35*, 112–117.

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. L. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *eLife*, *7*, 1–25. https://doi.org/10.7554/eLife.38677

van Moorselaar, D., Daneshtalab, N., & Slagter, H. A. (2021). Neural mechanisms underlying distractor inhibition on the basis of feature and/or spatial expectations. *Cortex*, *137*, 232–250. https://doi.org/10.1016/j.cortex.2021.01.010

van Moorselaar, D., Lampers, E., Cordesius, E., & Slagter, H. A. (2020). Neural mechanisms underlying expectation-dependent inhibition of distracting information. *eLife*, *9*, 1–26. https://doi.org/10.7554/ELIFE.61048

van Moorselaar, D., & Slagter, H. A. (2020). Inhibition in selective attention. *Annals of the New York Academy of Sciences*, *1464*(1), 204–221. https://doi.org/10.1111/nyas.14304

Vandierendonck, A. (2017). A comparison of methods to combine speed and accuracy measures of performance: A rejoinder on the binning procedure. *Behavior Research Methods*, *49*(2), 653–673. https://doi.org/10.3758/s13428-016-0721-5

Vatterott, D. B., & Vecera, S. P. (2012). Experience-dependent attentional tuning of distractor rejection. *Psychonomic Bulletin and Review*, *19*(5), 871–878. https://doi.org/10.3758/s13423-012-0280-4

Vickery, T. J., King, L.-W., & Jiang, Y. (2005). Setting up the target template in visual search. *Journal of Vision*, *5*(1), 8. https://doi.org/10.1167/5.1.8

Võ, M. L. H. (2021). The meaning and structure of scenes. *Vision Research*, *181*(October 2020), 10–20. https://doi.org/10.1016/j.visres.2020.11.003

Võ, M. L. H., & Henderson, J. M. (2010). The time course of initial scene processing for eye movement guidance in natural scene search. *Journal of Vision*, *10*(3), 1–13. https://doi.org/10.1167/10.3.14

Vossel, S., Mathys, C., Stephan, K. E., & Friston, K. J. (2015). Cortical coupling reflects Bayesian belief updating in the deployment of spatial attention. *Journal of Neuroscience*, *35*(33), 11532–11542. https://doi.org/10.1523/JNEUROSCI.1382-15.2015

Vossel, S., Thiel, C. M., & Fink, G. R. (2006). Cue validity modulates the neural correlates of covert endogenous orienting of attention in parietal and frontal cortex. *NeuroImage*, *32*(3), 1257–1264. https://doi.org/10.1016/j.neuroimage.2006.05.019

Wadia, V. S., Reed, C. M., Chung, J. M., Bateman, L. M., Mamelak, A. N., Rutishauser, U., & Tsao, D. Y. (2024). *A shared code for perceiving and imagining objects in human ventral temporal cortex* (p. 2024.10.05.616828). bioRxiv. https://doi.org/10.1101/2024.10.05.616828

Walther, A., Nili, H., Ejaz, N., Alink, A., Kriegeskorte, N., & Diedrichsen, J. (2016). Reliability of dissimilarity measures for multi-voxel pattern analysis. *NeuroImage*, *137*, 188–200. https://doi.org/10.1016/j.neuroimage.2015.12.012

Weber, S., Haynes, J., Christophel, T., Görgen, K., & Soch, J. (2024). Working memory signals in early visual cortex are present in weak and strong imagers. *Human Brain Mapping*, *45*(September 2023), e26590. https://doi.org/10.1002/hbm.26590

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005

Wiegand, I., Töllner, T., Dyrholm, M., Müller, H. J., Bundesen, C., & Finke, K. (2014). Neural correlates of age-related decline and compensation in visual attention capacity. *Neurobiology of Aging*, *35*(9), 2161–2173.

Williams, L. H., Wiegand, I., Lavelle, M., Wolfe, J. M., Fukuda, K., Peelen, M. V., & Drew, T. (2024). Electrophysiological Correlates of Visual Memory Search. *Journal of Cognitive Neuroscience*, 1–23. https://doi.org/10.1162/jocn_a_02256

Witkowski, P. P., & Geng, J. J. (2019). Learned feature variance is encoded in the target template and drives visual search. *Visual Cognition*, *27*(5–8), 487–501. https://doi.org/10.1080/13506285.2019.1645779

Witkowski, P. P., & Geng, J. J. (2022). Attentional priority is determined by predicted feature distributions. *Journal of Experimental Psychology: Human Perception and Performance*. https://doi.org/10.1037/xhp0001041

Wolfe, J. M. (1994). Guided Search 2. 0 A revised model of visual search. *Psychonomic Bulletin & Review*, *1*(2), 202–238.

Wolfe, J. M. (2012). Saved by a Log: How Do Humans Perform Hybrid Visual and Memory Search? *Psychological Science*, *23*(7), 698–703. https://doi.org/10.1177/0956797612443968

Wolfe, J. M. (2020). Visual Search: How Do We Find What We Are Looking For? *Annual Review of Vision Science*, *6*, 539–562. https://doi.org/10.1146/annurev-vision-091718

Wolfe, J. M. (2021). Guided Search 6.0: An updated model of visual search. *Psychonomic Bulletin and Review*, *28*(4), 1060–1092. https://doi.org/10.3758/s13423-020-01859-9

Wolfe, J. M., Alvarez, G. A., Rosenholtz, R., Kuzmova, Y. I., & Sherman, A. M. (2011). Visual search for arbitrary objects in real scenes. *Attention, Perception, & Psychophysics*, *73*(6), 1650–1671. https://doi.org/10.3758/s13414-011-0153-3

Wolfe, J. M., Cave, K. R., & Franzel, S. L. (1989). Guided Search: An Alternative to the Feature Integration Model for Visual Search. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 419–433. https://doi.org/10.1037/0096-1523.15.3.419

Wolfe, J. M., & Horowitz, T. S. (2004). What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience*, *5*(6), 495–501. https://doi.org/10.1038/nrn1411

Wolfe, J. M., & Horowitz, T. S. (2017). Five factors that guide attention in visual search. *Nature Human Behaviour*, *1*(3), 1–8. https://doi.org/10.1038/s41562-017-0058

Wolfe, J. M., Horowitz, T. S., Kenner, N., Hyle, M., & Vasan, N. (2004). How fast can you change your mind? The speed of top-down guidance in visual search. *Vision Research*, *44*(12), 1411–1426. https://doi.org/10.1016/j.visres.2003.11.024

Wolfe, J. M., Võ, M. L. H., Evans, K. K., & Greene, M. R. (2011). Visual search in scenes involves selective and nonselective pathways. *Trends in Cognitive Sciences*, *15*(2), 77–84. https://doi.org/10.1016/j.tics.2010.12.001

Wolff, M. J., Jochim, J., Akyürek, E. G., & Stokes, M. G. (2017). Dynamic hidden states underlying working-memory-guided behavior. *Nature Neuroscience*, *20*(6), 864–871. https://doi.org/10.1038/nn.4546

Won, B. Y., Haberman, J., Bliss-Moreau, E., & Geng, J. J. (2020). Flexible target templates improve visual search accuracy for faces depicting emotion. *Attention, Perception, and Psychophysics*, *82*(6), 2909–2923. https://doi.org/10.3758/s13414-019-01965-4

Woodman, G. F., & Luck, S. J. (2007). Do the Contents of Visual Working Memory Automatically Influence Attentional Selection During Visual Search? *Journal of Experimental Psychology. Human Perception and Performance*, *33*(2), 363–377. https://doi.org/10.1037/0096-1523.33.2.363

Wynn, J. S., Ryan, J. D., & Moscovitch, M. (2020). Effects of prior knowledge on active vision and memory in younger and older adults. *Journal of Experimental Psychology: General*, *149*(3), 518–529. https://doi.org/10.1037/xge0000657

Xie, S., Kaiser, D., & Cichy, R. M. (2020). Visual Imagery and Perception Share Neural Representations in the Alpha Frequency Band. *Current Biology*, *30*(13), 2621-2627.e5. https://doi.org/10.1016/j.cub.2020.04.074

Xu, L., Paffen, C. L. E., Van Der Stigchel, S., & Gayet, S. (2024). Statistical Learning Facilitates Access to Awareness. *Psychological Science*, *35*(9), 1035–1047.

Xu, Y. (2020). Revisit once more the sensory storage account of visual working memory. *Visual Cognition*, *28*(5–8), 433–446. https://doi.org/10.1080/13506285.2020.1818659

Xu, Y. (2023). Parietal-driven visual working memory representation in occipito-temporal cortex. *Current Biology*. https://doi.org/10.1016/j.cub.2023.08.080

Yeh, L. C., Thorat, S., & Peelen, M. V. (2024). Predicting Cued and Oddball Visual Search Performance from fMRI, MEG, and DNN Neural Representational Similarity. *Journal of Neuroscience*, *44*(12), 1–10. https://doi.org/10.1523/JNEUROSCI.1107-23.2024

Yu, Q., Teng, C., & Postle, B. R. (2020). Different states of priority recruit different neural representations in visual working memory. *PLoS Biology*, *18*(6), 1–21. https://doi.org/10.1371/journal.pbio.3000769

Yu, X., & Geng, J. J. (2019). The attentional template is shifted and asymmetrically sharpened by distractor context. *Journal of Experimental Psychology: Human Perception and Performance*, *45*(3), 336–353. https://doi.org/10.1037/xhp0000609

Yu, X., Hanks, T. D., & Geng, J. J. (2022). Attentional Guidance and Match Decisions Rely on Different Template Information During Visual Search. *Psychological Science*, *33*(1), 105–120. https://doi.org/10.1177/09567976211032225

Yu, X., Johal, S. K., & Geng, J. J. (2022). Visual search guidance uses coarser template information than target-match decisions. *Attention, Perception, and Psychophysics*, *84*(5), 1432–1445. https://doi.org/10.3758/s13414-022-02478-3

Yu, X., Rahim, R. A., & Geng, J. J. (2023). Task-adaptive changes to the target template in response to distractor context: Separability versus similarity. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001507

Yu, X., Zhou, Z., Becker, S. I., Boettcher, S. E. P., & Geng, J. J. (2023). Good-enough attentional guidance. *Trends in Cognitive Sciences*, *27*(4), 391–403. https://doi.org/10.1016/j.tics.2023.01.007

Zeman, A. (2024). Aphantasia and hyperphantasia: Exploring imagery vividness extremes. *Trends in Cognitive Sciences*, *28*(5), 467–480. https://doi.org/10.1016/j.tics.2024.02.007

Zhang, Z., & Carlisle, N. B. (2023). Explicit Attentional Goals Unlock Implicit Spatial Statistical Learning. *Journal of Experimental Psychology: General*, *152*(8), 2125.

Zhou, Z., & Geng, J. J. (2024). Learned associations serve as target proxies during difficult but not easy visual search. *Cognition*, *242*(October 2023), 105648. https://doi.org/10.1016/j.cognition.2023.105648

Zhou, Z., & Geng, J. J. (2025). Preparatory attentional templates in prefrontal and sensory cortex encode target-associated information. *eLife*, *14*. https://doi.org/10.7554/eLife.104041.1

Zoccolan, D., Cox, D. D., & DiCarlo, J. J. (2005). Multiple object response normalization in monkey inferotemporal cortex. *Journal of Neuroscience*, *25*(36), 8150–8164. https://doi.org/10.1523/JNEUROSCI.2058-05.2005

Zoccolan, D., Kouh, M., Poggio, T., & DiCarlo, J. J. (2007). Trade-Off between Object Selectivity and Tolerance in Monkey Inferotemporal Cortex. *The Journal of Neuroscience*, *27*(45), 12292–12307. https://doi.org/10.1523/jneurosci.1897-07.2007

Appendices

Appendix A: Nederlandse Samenvatting

Appendix B: English Summary

Appendix C: Deutsche Zusammenfassung

Appendix D: Acknowledgments

Appendix E: About the Author

Appendix F: List of Publications

Appendix G: Research Data Management

Appendix H: Donders Graduate School

Appendix A: Nederlandse Samenvatting

Onze dagelijkse omgeving is complex en vol met verschillende objecten: bijvoorbeeld drukke stadsstraten, volle treinstations, supermarkten, et cetera… Niet alle objecten in de omgeving kunnen in detail door ons visuele systeem worden verwerkt. Ze zijn ook niet allemaal relevant voor onze lopende doelen. Om de relevante objecten te vinden, zoeken we vaak en schijnbaar moeiteloos door zulke rommelige omgevingen; we kijken bijvoorbeeld uit naar aankomende fietsers voordat we een fietspad oversteken, we zoeken naar onze vermiste sleutels in onze kamer, of naar een appel in de supermarkt. Hoe vinden en verwerken we die objecten efficiënt terwijl we andere, irrelevante objecten negeren?

Eerder onderzoek heeft aangetoond dat visuele zoekopdrachten afhankelijk zijn van een interne (mentale) representatie van het doelobject – het aandachts-template. Visuele input wordt bij voorkeur verwerkt op basis van de overeenstemming met dit template. Een belangrijke ontdekking in het onderzoek naar menselijke aandacht was een mogelijke neurale basis voor deze aandachts-templates: na een aanwijzing die het volgende zoekdoel aangeeft, verhoogden neuronen in de visuele cortex die selectief waren voor doelkenmerken (bijvoorbeeld voorkeur voor rode of ronde objecten bij het zoeken naar een appel) hun activiteit al vóór de presentatie van de zoekomgeving en dus zonder visuele input. Deze voorbereidende activiteit kan fungeren als een *bias*-signaal, dat de verwerking van input die op het doel lijkt, bevoordeelt ten opzichte van andere niet-overeenkomende objecten.

In dit proefschrift heb ik de dynamische interactie tussen zulke interne representaties en de externe wereld onderzocht. Ten eerste zou de structuur van de omgeving onze aandachts-templates kunnen beïnvloeden. Hoewel complex, biedt onze dagelijkse omgeving een breed scala aan statistische regulariteiten die benut kunnen worden. Wanneer we bijvoorbeeld naar een appel in de keuken zoeken, weten we dat deze rood en rond is, maar we weten ook dat ze waarschijnlijk in een fruitschaal op het aanrecht liggen, die niet alleen appels maar misschien ook gele bananen bevat. De eerste twee empirische hoofdstukken van dit proefschrift onderzoeken of en hoe aandachts-templates en voorbereidende activiteit worden aangepast op basis van zulke omgevingsregulariteiten.

Leggen we de nadruk op verschillende kenmerken van hetzelfde object afhankelijk van hun voorspellingskracht in de huidige context? In Hoofdstuk 2 onderzochten we of het aandachts-template voor complexe, tweedimensionale vormen werd aangepast op basis van voorspelbare afleiders (niet-doelen in de zoekomgeving).

We voerden een reeks online gedragsexperimenten uit waarbij deelnemers werden aangezet om te zoeken naar vormen binnen twee voorspellende contexten. In elke context was de doelvorm uniek en onderscheidde zich van de afleiders in één kenmerk: het was ofwel het enige horizontale/verticale of het enige ronde/vierkante object in de weergave. In meerdere experimenten lieten we zien dat de meer voorspellende kenmerken in de huidige context de voorkeur kreeg boven de minder voorspellende kenmerken. Dit werd betrouwbaar gevonden in geblokte contexten, waarin de voorspellende dimensie herhaald werd over meerdere proeven. Bovendien testten we of het template ook kon worden aangepast op basis van contextuele aanwijzingen per trial. We ontdekten dat dit het geval was, maar alleen wanneer de twee afleidende contexten ruimtelijk gescheiden waren en verschillende achtergrondkleuren hadden. Deze resultaten ondersteunen de opkomende visie dat het aandachts-template geen statische en waarheidsgetrouwe representatie is van doelkenmerken, maar een representatie die flexibel kan worden aangepast voor efficiënte zoekopdrachten. Belangrijk is dat onze resultaten aantonen dat we inderdaad flexibel verschillende dimensies van een eenheidsvorm kunnen benadrukken, en dat dit zowel kan worden gedaan op basis van de recente selectiegeschiedenis als op basis van contextuele aanwijzingen per trial.

Aandacht in omgevingen in de echte wereld wordt vaak geleid door niet-doelkenmerken, inclusief grote doel-geassocieerde objecten die gemakkelijk te vinden zijn en die de locatie van het doel kunnen voorspellen – zogenaamde "anker"-objecten. Bijvoorbeeld, wanneer we zoeken naar een pen in een kantoor, zouden we eerst kunnen kijken naar het bureau, dat makkelijk te zien is en waar de pen zich waarschijnlijk zal bevinden. In Hoofdstuk 3 onderzochten we voorbereidende neurale activiteit in context-geleide zoekopdrachten, met behulp van fMRI en gelijktijdige oogbewegingstracking. Deelnemers zochten naar doelobjecten in twee contexten die geassocieerd waren met verschillende door aandacht geleide ankerobjecten. We vonden dat de voorbereidende activiteit het relevante geleidingsobject in de huidige context weerspiegelde. Daarentegen was er geen bewijs voor een voorbereidende representatie van het zoekdoel. Door onderscheid te maken tussen kenmerken die relevant waren voor het geleiden van aandacht en kenmerken die relevant waren voor doelidentificatie, toonden we aan dat de voorbereidende activiteit voornamelijk de eerste ondersteunt, en een flexibele (contextafhankelijke) en intern gegenereerde geleidings-template reflecteert in plaats van de kenmerken van het aangegeven doel. Het voorbereidende template kan dus niet alleen een enigszins vervormde versie van het doel weerspiegelen, maar ook volledig andere objecten vertegenwoordigen. Dit levert bewijs voor het eerder veronderstelde onderscheid tussen doel- en geleidings-templates en werpt licht op de functionele rol van voorbereidende activiteit in context-geleide zoekopdrachten.

Een andere relevante vraag is hoe intern gegenereerde representaties in de visuele cortex tegelijkertijd bestaan met representaties van externe objecten. Tijdens veel alledaagse situaties, niet alleen tijdens zoekopdrachten maar ook wanneer we objecten herinneren of visueel verbeelden, zijn intern gegenereerde representaties gelijktijdig aanwezig met sensorische input. Bijvoorbeeld, we kunnen een luisterboek horen en de beschreven personages en plaatsen visueel verbeelden, terwijl we tegelijkertijd waarnemen dat een kat onze kamer binnenkomt. Dit creëert potentiële concurrentie tussen aparte externe en interne representaties in de visuele cortex, vooral wanneer beide gebruik maken van overlappende neurale middelen. Als dit het geval is, kunnen we minder van onze omgeving waarnemen wanneer we tegelijkertijd een vergelijkbaar object verbeelden. Deze vraag is het onderwerp van het laatste experimentele hoofdstuk. In Hoofdstuk 4 gebruikten we EEG om de interactie tussen verbeelding en perceptie te bestuderen. We ontdekten dat het verbeelden van een object uit dezelfde categorie de perceptuele verwerking van externe objecten verstoorde, in vergelijking met objecten uit een andere categorie. Deze bevinding levert bewijs voor inhoud-specifieke concurrentie tussen representaties van verbeelde en waargenomen objecten. Deze concurrentie was specifiek voor objecten uit dezelfde categorie, zodat deze afwezig was voor objecten die binnen hetzelfde domein (dieren of gebouwen) vielen maar uit een andere categorie komen (bijv. het zien van een zoogdier en het verbeelden van een vogel), waarschijnlijk omdat deze representaties minder vergelijkbaar waren. Terwijl dit proefschrift, en onderzoek in het algemeen, zich grotendeels heeft gericht op het faciliterende effect van interne representaties, vooral tijdens visuele zoekopdrachten, leveren we hier ook neurale bewijs voor het tegenovergestelde effect (d.w.z. interferentie) in een verbeeldingstaak. Deze studie biedt nieuwe inzichten in de interacties tussen top-down en bottom-up signalen in de visuele cortex.

Al met al biedt dit proefschrift nieuwe inzichten in de wederzijdse interacties tussen interne representaties in de visuele cortex en de externe wereld. In plaats van statisch en noodzakelijkerwijs waarheidsgetrouw te zijn, kunnen deze interne representaties, weerspiegeld in inhoud-specifieke top-down activatie in de visuele cortex, flexibel en contextafhankelijk zijn, en kunnen ze, onder bepaalde omstandigheden, niet alleen de perceptuele verwerking vergemakkelijken, maar er ook mee concurreren.

(translated with the help of Charlotte de Blecourt and ChatGPT)

Appendix B: English Summary

Our typical environments, e.g., busy city streets, crowded train stations or supermarkets, are complex and full of different objects. Not all of the objects in these scenes can be processed in detail by our visual system, nor are they all relevant to our current goals. To find those relevant objects, we search through such cluttered environments frequently and with apparent ease; for example, we look out for approaching bikes before crossing a bike path, we search for our missing keys in our room, or for an apple in the supermarket. How do we find and process those objects efficiently while ignoring other, irrelevant ones?

Previous research has shown that visual search relies on an internal (mental) representation of the target - the attentional template. Visual input is preferentially processed based on its match with this template. A key discovery in attention research was a potential neural basis for these attentional templates: after a cue indicating the following search target, neurons in visual cortex that were selective to target features (e.g., preferring red or round objects when searching for an apple) increased their activity already before the search scene was presented and therefore in the absence of visual input. This preparatory activity could act as a biasing signal, favouring processing of input resembling the target over other non-matching objects.

In this thesis, I have investigated the dynamic interplay between such internal representations and the external world. One important aspect is that the structure of the environment could shape our attentional templates. While complex, our daily surroundings offer a wide range of statistical regularities that could be exploited for efficient search. For instance, when looking for an apple in the kitchen, we know it is red and round, but also know they are likely to be found in a fruit basket, placed on the countertop, containing not only apples but perhaps also yellow bananas. The first two empirical chapters of this thesis investigate whether and how attentional templates and preparatory activity are adapted based on such environmental regularities.

Do we emphasise different feature-dimensions of the same object depending on whether they are diagnostic in the current context? In Chapter 2, we asked whether the attentional template for complex, two-dimensional shapes was adapted based on predictable distractors (non-targets in the search scene). We conducted a series of online behavioural experiments in which participants were cued to search for shape targets within two probabilistic contexts. In each of these contexts, the target shape was unique and distinct from distractors in one feature dimension: it was either being the only horizontal/vertical or the only round/square object in the display. Across multiple experiments, we showed that the more diagnostic dimension in the current context was emphasized over the less diagnostic one. This was reliably found in blocked contexts, in which the diagnostic dimension repeated across multiple trials. In addition, we tested whether the template could also be adapted based on trial-by-trial context cues. We found that this was the case, but only when the two distractor contexts were spatially separated and had different background colours. These results support the emerging view that the attentional template is not a static and veridical representation of target features, but a representation that can be flexibly adapted for efficient search. Importantly, our results show that we can indeed flexibly emphasise distinct dimensions of a unitary shape, and that this can be done both based on recent selection history as well as trial-wise context cues.

Attentional guidance in real-world scenes is often based on non-target features, including large target-associated objects that are easy to find and are predictive of the target's location – so-called "anchor" objects. For instance, looking for a pen in the office, we may first look at the easily visible office desk on which we expect it to be. In Chapter 3, we investigated preparatory neural activity in context-guided search, using fMRI and concurrent eye tracking. Participants searched for target objects that were associated with different attention-guiding anchor objects across two contexts. We found that preparatory activity reflected the relevant guiding object in the current context. By contrast, there was no evidence for a preparatory target representation. By dissociating between features that were relevant for attentional guidance and features that were relevant for target identification, we showed that preparatory activity primarily supports the former, reflecting a flexible (context-dependent) and internally-generated guiding template rather than the features of the cued target. The preparatory template can thus not only reflect a somewhat distorted version of the target, but can also represent entirely different objects instead. This provides evidence for the previously hypothesized distinction between target and guiding templates and sheds light on the functional role of preparatory activity in context-guided search.

Another relevant question is how internally-generated representations co-exist in visual cortex with representations of external objects. During many everyday situations, not only during search but also when we remember, or visually imagine objects, internally generated representations are concurrently present with sensory input. For instance, we may listen to an audiobook and visually imagine the described characters and places, while also perceiving that a cat has entered our room. This creates potential competition between separate external and internal representations within visual cortex, especially when both draw on overlapping neural resources. If this is the case, we may perceive less of our environments when simultaneously imagining a similar object. This question is the focus of the final experimental chapter. In Chapter 4, we used EEG to study the interaction between imagery and perception. We found that imagining an object of the same category impaired perceptual processing of external objects, as compared to objects of a different category. This finding provides evidence for content-specific competition between representations of imagined and perceived objects. This competition was specific to same-category objects, such that it was absent for objects that were within the same domain (animals or buildings) but from a different category (e.g., seeing a mammal and imagining a bird), likely because these representations were less similar. While this thesis, and research overall, has largely focused on the facilitatory effect of internal representations, especially during visual search, here we also provide neural evidence for the opposite effect (i.e., interference) in an imagery task. This study provides new insights into the interactions of top-down and bottom-up signals within visual cortex.

Overall, this thesis provides new insights into the bidirectional interactions between internal representations in visual cortex and the external world. Rather than being static and necessarily veridical, these internal representations can be flexible and context-dependent, and can, under some circumstances, not only facilitate perceptual processing but also compete with it.

Appendix C: Deutsche Zusammenfassung

Unsere alltäglichen Umgebungen, wie z.B. belebte Straßen in der Stadt, Bahnhöfe oder Supermärkte, sind komplex und beinhalten eine Vielzahl verschiedener Objekte. Weder kann unser visuelles System jedoch all diese Objekte detailliert verarbeiten, noch sind alle diese Objekte, relevant für unsere momentanen Ziele. Um relevante Objekte zu finden, suchen wir, oft und scheinbar mühelos, durch diese überfüllten Umgebungen: wir schauen nach herannahenden Fahrrädern beim Überqueren des Radweges, suchen in unserem Zimmer nach unseren verschwundenen Schlüsseln oder im Supermarkt nach einem Apfel. Wie gelingt es uns, effizient relevante Objekte zu finden und dabei gleichzeitig andere, irrelevante zu ignorieren?

Bisherige Forschungen haben ergeben, dass für die visuelle Suche eine interne (mentale) Repräsentation des Zielobjektes genutzt wird – die sogenannte Aufmerksamkeitsschablone (attentional template). Basierend auf der Übereinstimmung mit dieser Schablone werden Teile des visuellen Inputs bevorzugt verarbeitet. Eine bedeutende Entdeckung in der Aufmerksamkeitsforschung war die Identifizierung einer potenziellen neuronalen Basis für solche Aufmerksamkeitsschablonen: kurz nachdem ein Hinweis, der das nächste Zielobjekt angab, erschien, begannen Neuronen im visuellen Kortex, die selektiv für Merkmale dieses Zielobjektes waren (z.B. eine Präferenz für rote oder runde Objekte hatten wenn nach einem Apfel gesucht werden sollte) ihre Aktivität zu erhöhen, bereits vor dem Erscheinen einer Suchszene, und damit in Abwesenheit von visuellem Input. Diese ‚vorbereitende Aktivität' (preparatory activity) könnte als Biassignal fungieren, um die Weiterverarbeitung von Input der dem Zielobjekt ähnlich ist gegenüber anderem, nicht übereinstimmenden, zu bevorzugen.

In dieser Dissertation habe ich die dynamischen Interaktionen zwischen solch internen Repräsentationen und der externen Welt untersucht. Ein wichtiger Aspekt ist, dass die Struktur der Umgebung unsere Aufmerksamkeitsschablonen beeinflussen könnte. Zwar sind unsere alltäglichen Umgebungen komplex, aber auch voller statischer Regelmäßigkeiten die für eine effizientere Suche genutzt werden könnten. So wissen wir zum Beispiel wenn wir nach einem Apfel in der Küche suchen, nicht nur dass er rot und rund ist, sondern auch dass Äpfel häufig in einer Obstschale auf der Küchentheke liegen, möglicherweise neben ein paar gelben Bananen. In den ersten zwei empirischen Kapiteln dieser Arbeit erforschen wir ob und wie Aufmerksamkeitsschablonen an solche Regelmäßigkeiten in unserer Umgebung angepasst werden.

Heben wir, wenn wir in verschiedenen Kontexten nach demselben Objekt suchen, unterschiedliche Merkmalsdimensionen hervor, je nachdem welche im jetzigen Kontext diagnostisch sind? In Kapitel 2 untersuchten wir, ob die Aufmerksamkeitsschablone für komplexe, zweidimensionale Formen an vorhersehbare Distraktoren (nicht-Zielobjekte in der Suchszene) angepasst werden kann. In einer Reihe von Online-Verhaltensexperimenten waren Probanden angewiesen nach bestimmten Ziel-formen in zwei probabilistischen Kontexten zu suchen. In jeder der beiden Kontexte war die Ziel-form einzigartig und unterschied sich von anderen Distraktoren in einer Merkmalsdimension: entweder war sie die einzige horizontale/vertikale oder runde/eckige Form im Suchdisplay. In mehreren Experimenten wiesen wir nach, dass die diagnostischere Merkmalsdimension gegenüber der weniger diagnostischen bevorzugt wurde. Dies war über mehrere Experimente hinweg der Fall, wenn der Distraktor-kontext geblockt war, und die diagnostische Dimension über mehrere Versuchsdurchgänge (trials) dieselbe blieb. Zusätzlich testeten wir ob die Aufmerksamkeitsschablone auch anhand von Kontext-hinweisen zu Beginn jedes Versuchsdurchganges angepasst werden konnte. Dies war möglich, allerdings nur wenn beide Kontexte durch räumliche Trennung und unterschiedliche Hintergrundfarben gekennzeichnet waren. Diese Ergebnisse stützen die aufkommende Sichtweise, dass eine Aufmerksamkeitsschablone keine statische und wahrheitsgetreue Repräsentation der Merkmale des Zielobjektes ist, sondern für eine effiziente Suche angepasst werden kann. Insbesondere wiesen wir nach, dass wir in der Tat flexibel verschiedene Dimensionen einer einheitlichen Form hervorheben können, basierend sowohl auf Grund der unmittelbaren vorhergegangen Versuchsdurchgänge (selection history), als auch durch Kontext-hinweise zu Beginn jedes Versuchsdurchgangs.

In realistischen Suchszenen wird unsere Aufmerksamkeit häufig durch Eigenschaften von nicht-Zielobjekten gelenkt, z.B. durch sogenannte Anker-Objekte (anchor objects): große Objekte die einfach zu finden sind, und genutzt werden können um vorauszusagen *wo* sich das Zielobjekt befindet. Wenn wir zum Beispiel nach einem Stift suchen, schauen wir wahrscheinlich zuerst auf den leicht sichtbaren Schreibtisch, auf dem wir den Stift erwarten würden. In Kapitel 3 haben wir, mit funktionalem MRT und gleichzeitigen Eye-tracking, vorbereitende Aktivität in solch kontext-geführter Suche untersucht. Diese spiegelte das relevante aufmerksamkeits-lenkende Anker-Objekt im jetzigen Kontext wieder, jedoch nicht das eigentliche Zielobjekt. In dem wir zwischen Merkmalen zur Aufmerksamkeitslenkung und zur Identifizierung des Zielobjekts unterschieden haben, konnten wir zeigen, dass vorbereitende Aktivität hauptsächlich der Aufmerksamkeitslenkung dient und eine (kontextabhängiges) und intern generierte Führungs-Schablone (guiding template) darstellt. Dies bedeutet dass vorbereitende Aufmerksamkeitsschablonen nicht nur eine leicht verzerrte Version des Zielobjektes, sondern sogar gänzlich andere (nicht-Ziel-)Objekte darstellen können. Unsere Ergebnisse liefern eine erste empirische Grundlage zur bisher theoretischen Unterscheidung zwischen Schablonen zur Aufmerksamkeitslenkung und zur Identifikation des Zielobjektes, und klären welche funktionale Rolle vorbereitende Aktivität bei der kontext-geführten Suche spielt.

Eine weitere relevante Frage ist, wie intern generierte Repräsentationen gemeinsam mit Repräsentationen externer Objekten im visuellen Kortex koexistieren. In vielen Alltagssituationen, nicht nur während wir suchen, sondern auch wenn wir uns an Objekte erinnern oder sie uns vor unserem ‚inneren Auge' vorstellen, sind diese internen Repräsentationen und der eingehende sensorischer Input zeitgleich im visuellen Kortex präsent. Zum Beispiel, wenn wir ein Hörbuch hören, uns dabei die beschriebenen Charaktere und Orte lebhaft bildlich vorstellen, und dabei gleichzeitig feststellen dass eine Katze unser Zimmer betreten hat. Dies könnte potentiell dazu führen, dass intern- und extern-generierte Signale im visuellen Kortex konkurrieren, besonders wenn beide überlappende neuronale Ressourcen beanspruchen. Wäre dies der Fall, würden wir weniger von unserer Umgebung wahrnehmen, wenn wir uns gleichzeitig ein ähnliches Objekt lebhaft bildlich vorstellen. Diese Frage untersuchten wir im letzten empirischen Kapitel (Kapitel 4). Wir nutzten EEG um die Interaktion zwischen diesen mentalen Bildern und unserer Wahrnehmung zu testen. Unsere Ergebnisse zeigten, dass die Wahrnehmungsverarbeitung eines Objektes gestört war, wenn Probanden sich gleichzeitig ein Objekt derselben Kategorie vorstellten. Dies weist auf inhaltsspezifische Konkurrenz zwischen Wahrnehmung und bildlicher Vorstellung hin. Diese Interferenz war spezifisch für Objekte derselben Kategorie, d.h. wir fanden keine Interferenz wenn die präsentierten und bildlich vorgestellten Objekte derselben Domäne (Tiere oder Gebäude) kamen, aber unterschiedlichen Kategorien angehörten (z.B. wenn wir ein Säugetier sehen, während wir uns einen Vogel vorstellen), wahrscheinlich da sich diese Repräsentation stärker voneinander unterschieden. Obwohl der Fokus dieser Dissertation und der Forschung im Allgemeinen darauf liegt, wie interne Repräsentation die Wahrnehmung unterstützen, liefern wir hier empirische (neuronale) Evidenz für den gegenteiligen Effekt. Diese Studie ermöglicht neue Einblicke in die Interaktionen zwischen intern und extern generierten Signalen im visuellen Kortex.

Insgesamt bietet diese Dissertation neue Erkenntnisse zur bidirektionalen Interaktion zwischen internen Repräsentationen im visuellen Kortex und der externen Welt. Statt statisch und notwendigerweise wahrhaftgetreu zu sein, können diese internen Repräsentationen flexibel und kontext-abhängig sein, und die Wahrnehmungsverarbeitung nicht nur unterstützen, sondern, unter bestimmten Umständen, auch mit ihr konkurrieren.

Appendix D: Acknowledgments

Doing a PhD is never an entirely smooth ride, with many ups and downs on the way. My own experience was no exception there, yet looking back I was often very lucky with how things turned out, and had many people around me who have supported and guided me through the years. I am very grateful for every one of them!

Marius, you were the real MVP of this PhD. Thank you so much for giving me the chance to (at least partially) answer the questions I had on my mind since sitting in your lecture as a Master student, your trust in me and continuous guidance throughout the past years. In each and every meeting, you have impressed me with your clarity of thought, your dedication to good science and your ability to remember and grasp all the details of my projects. You gently challenged and encouraged me to reach further, to trust myself and develop my own ideas, but also reassured me when I was not meeting my own standards, or things did not go according to plan. Through all the successful moments, some setbacks, doubts, and frustrating results, I knew that I could count on you!

Floris, thank you for agreeing to be part of my projects! I value your opinion, kind and relaxed approach to science and sometimes critical questions that made me rethink my assumptions or the way I explained things. Even though we sometimes did not see each other for a while, thank you for making the time to e.g., be there for the first MRI scan of my guiding templates project and for inviting me to join your statistical learning club!

Thanks to all of my incredibly kind, smart, weird and funny labmates from all over the world! Being part of **Peelenlab** meant I did not only have colleagues to discuss science, papers or the latest frustrating experiments, but friends I got to share so many more experiences, trips, celebrations and lots of food with! Thanks for all the whiskey evenings, karaoke nights, concerts, dumplings, pizzas, Chinese festivals, Sinterklaas and Indian wedding craziness! **Qiu**, you have been an essential companion throughout the last four years, to debate, travel, laugh, dance and even organize a conference with! Thank you for always welcoming me at KK25! I admire your creative talent and unstoppable determination in everything you do. Now that you are also almost done, I will miss all of our scientific and non-scientific discussions (at least most of them, haha), but have no doubt an exciting future awaits you! **Marco**, thanks for keeping all of us well-fed and mostly sane! Mille grazie for your mental support and having someone to rant or laugh with while finishing my PhD, transitioning into postdoc life and generally dealing with both

the good and the more frustrating parts of research. **Charlotte,** you have been with me since the start of this journey, also as a travel and hiking companion and soon as paranymph. Thank you for regularly checking in on me at the beginning of my PhD, helping me get started with my eye tracking and EEG projects, having an open ear and always being there for a dose of dark humour and more serious discussions! **Songyun,** thank you for often making my day when coming to the office! May your future be bright, without too many pigeons in your way! **Lu,** my academic older sister from Taiwan, who shares my interest in attentional templates, whiskey, Christmas cookies and much more (including height). I am glad we crossed paths in Peelenlab and that I can see you turn into an academic superstar now! I still hope we'll get to collaborate on a project one day! **Sush,** you made me feel welcome in Peelen lab as a Master student, and have always tried to challenge me in a good way with your endless curiosity and openness! **Giacomo,** thanks for being my Blender guru and dance buddy, helping me set up virtual living rooms and rock all the dancefloors from Nijmegen, Egmond an Zee, Aberdeen and finally Karad! **Surya,** I really lucked out by having you as a Master thesis supervisor! Merci for introducing me to the world of attentional templates, preparatory activity, fMRI and your continuous support! Thank you also to **Alexandra, Gen, Simen, Chuanji, Yuanfang, Lydia, Eelke, Linlin** and all the other Peelenlab members over the years!

I also owe a lot too my students, especially **Daan, Maria** and **Cara** who have helped me create stimuli, collected data and discussed ideas with me!

All the rest of DCC and Donders, especially my fellow PhDs and Postdocs including **Anne, Bingjie, Burcu, Charlotte, Claire, Helena, Eva, Feyi, Floortje, Johannes, Mara, Marlijn, Naomi, Orhun, Pierre, Pierre-Pascal, Siddharth, Tilman** as well as the DCC and DCCN support staff including **Jiska, Karin, Maarten, Pascal, Paul, Pauline, Philip and Vanessa.**

Of course, none of my research would have been possible without the help of many, many **participants** who agreed to take part in all of my studies!

Far beyond research, I was glad to have many old and new friends supporting me in the past four years!

Lisa and Boyd, thanks for making Covid times not only bearable, but sometimes even fun as old friends and flatmates in 360. Some of my favourite Covid memories are our quarter-past-curfew walks and the celebration of our doomed trip to Athens (which we luckily still went on with two years delay). Thanks Boyd for pushing me to run two marathons (which helped me keep me sanity here and there) and Lisa for being my close and constant companion in these early Covid and PhD times, and all the many tears of laughter over the years!

Yingjie, I have so many memories of all the things we did together over the last six years! Thank you for being an awesome friend, classmate, shorttime flatmate and now paranymph! I am glad I was able the highs and lows of PhD life with you, or to run, hike, surf, swim, SUP, skate or boulder together to distract us from them!

Lina, wir haben seit unserem ersten Treffen auf dem Schulhof vor über 20 Jahren und über verschiedene Länder und Kontinente hinweg so viel gemeinsam erlebt, und dürfen uns nun bald beide „Frau Doktor“ nennen. Wer hätte das damals gedacht? Ich bin unglaublich dankbar damals so eine treue Freundin gefunden zu haben, und das wir trotz unterschiedlichen Lebenswegen und weniger häufigen Treffen doch immer wieder Parallelen finden. Auf hoffentlich noch viele, viele weitere Jahre, ich bin gespannt wo es uns beide noch hinzieht!

Eva, wir kennen uns auch fast schon ewig. Danke für die gemeinsamen Auszeiten am Meer oder in den Bergen und regelmäßige Nachrichten!

Laura, murakoze cyane für die gemeinsamen Trails und Telefonate. Karibu Nijmegen!

Vince and Layla, thanks for all the meals, discussions and laughs we shared in 360!

Max, thanks for being my hiking buddy during the last phase of my PhD. I really enjoyed this time away from the laptop and the company!

Kirsten, danke für die inzwischen 11 Jahre Freundschaft, und Postkarten, Anrufe, gemeinsame Konferenzbesuche und Irland-Roadtrips in den letzten Jahren! Ich hoffe immer noch ich schaffe es bald auch nach London!

Danke an **Carmen, Ela, Wiebke, Nadia, Torsten, Noah, Andreas, Gesine, Charlotte, Sophie** und die vielen, vielen weiteren Menschen die mich auf meinem bisherigen Weg begleitet und unterstützt haben haben!

Und meine Familie. **Mama,** danke dass du immer an mich geglaubt hast und für deine bewundernswerte Stärke und Einsatz! Du hast mich von klein auf tatkräftig unterstützt meinen eigenen Weg zu gehen und dich immer für mich eingesetzt, trotz aller Widrigkeiten. Ohne dich hätte ich es nie so weit bringen können. **Aurèle**, du bist der beste kleine/große Bruder den ich mir wünschen kann. Danke für all deinen Humor, Unterstützung, Memes und Gespräche die wohl nur wir beide führen können! **Ulli,** Beutepapa, was ich von dir gelernt habe hat mich die letzten Jahre weiterhin begleitet, und ich würde diesen Meilenstein so gerne noch gemeinsam mit dir feiern können! Danke für alles!

Appendix E: About the Author

Maëlle Lerebourg was born in Berlin on 27th August 1995. Soon, she moved to Düsseldorf, where she later attended the Luisengymnasium, and very briefly Whangaparaoa High School in Whangaparaoa (New Zealand). During high school, she followed specializations in Biology and French and earned the French-German high school diploma Abi-Bac. During this time, she also heard for the first time that Neuroscience was a potential university degree, and was immediately intrigued by the idea of studying the brain. Before starting university, however, she decided to complete a voluntary service and spent a year working as assistant teacher in two secondary schools in Rwanda.

In 2014, she began studying Cognitive Science in Osnabrück (Germany). While she wasn't entirely sure what this was yet, this sounded very interesting and promised to include a broad range of topics from Artificial Intelligence, Computer Science, Neuroinformatics, Linguistics, Philosophy, to Maths and Psychology. While initially more drawn to Philosophy, Linguistics and Psychology, she quickly discovered she also enjoyed coding, and her favourite course became Action & Cognition, which focused on visual perception and sensorimotor control. During this time, she also became a tutor and workgroup teacher for different computer science, psychology and neuroscience courses. As part of her studies, she spent six months as an intern at the Center on Autobiographical Memory Research (Con Amore) in Aarhus (Denmark), where she investigated the development of children's visuospatial memory skills using eyetracking. Back in Osnabrück, she completed her bachelor's thesis in the Neurobiopsychology research group, testing the face-specificity of the N170 ERP component using EEG, under the supervision of Dr. Anna Lisa Gert and Prof. Peter König. This project was her first step into investigating naturalistic vision.

She then decided to specialize further and moved to Nijmegen to study Cognitive Neuroscience in 2018. In line with her previous interests, she chose the Perception, Action and Control track track. Besides her studies, she played an active role in the student community, as editor of the CNS student journal that published student's theses, and in the study association Dondrite, as chair of the committee organizing the annual Synapsium student conference and later as member of the travel committee. In 2019, she joined the Visual Cognitive Neuroscience group of Prof. Marius Peelen for her internship. There she used fMRI to study how preparatory attentional templates are rescaled to account for depth-related changes in retinal size, under supervision of Dr. Surya Gayet. She graduated summa cum laude and began her PhD in October 2020.

Besides her PhD research, she was also engaged in supervising Bachelor and Master students, worked as a teaching assistant and workgroup teacher for different AI and Psychology Bachelor Courses (Brain & Cognition I, Academic Skills, Academic & Professional Skills) and helped to organize the annual Donders Discussions PhD conference.

Currently, she is a postdoc in the visual cognitive neuroscience group, studying competitive interactions between mental imagery and perception.

Appendix F: List of Publications

Thesis Chapters

Lerebourg, M., de Lange, F. P., & Peelen, M. V. (2024). Attentional Guidance Through Object Associations in Visual Cortex. *Science Advances*.

Lerebourg, M., de Lange, F. P., & Peelen, M. V. (2023). Expected Distractor Context Biases the Attentional Template for Target Shapes. *Journal of Experimental Psychology. Human Perception and Performance*.

Other publications

Krøjgaard, P., Sonne, T., **Lerebourg, M.**, Lambek, R., & Kingo, O. S. (2019). Eight-year-olds, but not six-year-olds, perform just as well as adults when playing Concentration: Resolving the enigma?. *Consciousness and Cognition*, *69*, 81-94.

Appendix G: Research Data Management

This research followed the applicable laws and ethical guidelines. Research Data Management was conducted according to the FAIR principles. The paragraphs below specify in detail how this was achieved.

Ethics

This thesis is based on the results of human studies, which were conducted in accordance with the principles of the Declaration of Helsinki. The Ethical Committee of the faculty of Social Sciences (ECSS) has given a positive advice to conduct the studies reported in Chapter 2 and 4 to the Dean of the Faculty, who formally approved the conduct of these studies (ECSW2017-2306-517 and ECSW-LT-2022-9-9-25489 for Chapter 2, ECSW-2022-079 for Chapter 4). The study reported in chapter 3 was conducted under the ethical approval of the regional Commission for Research involving human subjects (CMO Arnhem Nijmegen, "CMP 2014/288). This research was partially supported by the European Research council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 725970) awarded to M.V. Peelen.

Findable Accessible

The table below details where the data and research documentation for each chapter can be found on the Donders Repository (DR). All data archived as a Data Sharing Collection remain available for at least 10 years after termination of the studies.

Chapter	DAC	RDC	DSC	DSC License	OSF
2	2021.00036_311	-	2021.00056_367	CC-BY-4.0	https://osf.io/tgu8r/
3	2022.00080_955	-	2022.00080_213	CC0-1.0	-
4	2024.00029_487	-	-	-	-

DAC = Data Acquisition Collection, RDC = Research Documentation Collection, DSC = Data Sharing Collection

Informed consent was obtained on paper following the Centre procedure. The forms are archived in the central archive of the Centre for 10 years after termination of the studies.

Interoperable, Reusable

The raw data are stored in the DAC in their original form. For the DSC long-lived file formats (e.g. .csv, .nii, .png, .m, .edf) have been used ensuring that data remains usable in the future. Results are reproducible by providing a description of the experimental setup, (anonymised) raw data, analysis scripts or pipelines (DAC and DSC).

Privacy

The privacy of the participants in this thesis has been warranted using random individual subject codes. A pseudonymization key linked this random code with the personal data. This pseudonymisation key was stored on a network drive that was only accessible to members of the project who needed access to it because of their role within the project. The pseudonymisation key was stored separately from the research data. The pseudonymization keys of chapter 2 and3 were destroyed within one month after finalization of these projects. The key of chapter 4 is still stored on a dedicated restricted network drive and will be destroyed within one month after finalization. Data in chapters 2 and 4 are not identifiable and shared without restrictions. Publicly shared MRI-data of chapter 3 in the DSC is defaced.

Appendix H: Donders Graduate School

For a successful research Institute, it is vital to train the next generation of scientists. To achieve this goal, the Donders Institute for Brain, Cognition and Behaviour established the Donders Graduate School in 2009. The mission of the Donders Graduate School is to guide our graduates to become skilled academics who are equipped for a wide range of professions. To achieve this, we do our utmost to ensure that our PhD candidates receive support and supervision of the highest quality.

Since 2009, the Donders Graduate School has grown into a vibrant community of highly talented national and international PhD candidates, with over 500 PhD candidates enrolled. Their backgrounds cover a wide range of disciplines, from physics to psychology, medicine to psycholinguistics, and biology to artificial intelligence. Similarly, their interdisciplinary research covers genetic, molecular, and cellular processes at one end and computational, system-level neuroscience with cognitive and behavioural analysis at the other end. We ask all PhD candidates within the Donders Graduate School to publish their PhD thesis in de Donders Thesis Series. This series currently includes over 600 PhD theses from our PhD graduates and thereby provides a comprehensive overview of the diverse types of research performed at the Donders Institute. A complete overview of the Donders Thesis Series can be found on our website: https://www.ru.nl/donders/donders-series

The Donders Graduate School tracks the careers of our PhD graduates carefully. In general, the PhD graduates end up at high-quality positions in different sectors, for a complete overview see https://www.ru.nl/donders/destination-our-former-phd. A large proportion of our PhD alumni continue in academia (>50%). Most of them first work as a postdoc before growing into more senior research positions. They work at top institutes worldwide, such as University of Oxford, University of Cambridge, Stanford University, Princeton University, UCL London, MPI Leipzig, Karolinska Institute, UC Berkeley, EPFL Lausanne, and many others. In addition, a large group of PhD graduates continue in clinical positions, sometimes combining it with academic research. Clinical positions can be divided into medical doctors, for instance, in genetics, geriatrics, psychiatry, or neurology, and in psychologists, for instance as healthcare psychologist, clinical neuropsychologist, or clinical psychologist. Furthermore, there are PhD graduates who continue to work as researchers outside academia, for instance at non-profit or government organizations, or in pharmaceutical companies. There are also PhD graduates who work in education, such as teachers in high school, or as lecturers in higher education. Others continue in a wide range of positions, such as policy advisors, project managers, consultants, data scientists, web- or software developers, business owners, regulatory affairs specialists, engineers, managers, or IT architects. As such, the career paths of Donders PhD graduates span a broad range of sectors and professions, but the common factor is that they almost all have become successful professionals.

For more information on the Donders Graduate School, as well as past and upcoming defences please visit: http://www.ru.nl/donders/graduate-school/phd/

9 789465 151519 >

DONDERS
INSTITUTE

Radboud University

Radboudumc