Preparatory activity during visual search reflects attention-guiding objects rather than search targets

Maëlle Lerebourg [1], Floris P. de Lange [1], & Marius V. Peelen [1]

Donders Institute for Brain, Cognition and Behaviour, Radboud University

Email: maelle.lerebourg@donders.ru.nl

Abstract

Efficient behavior requires the rapid attentional selection of task-relevant objects. Previous research has shown that target-selective neurons in visual cortex increase their baseline firing rate when participants are cued to search for a target object. Such preparatory activity represents a key finding for theories of visual search, as it may reflect a top-down bias that guides spatial attention, favoring processing of target-matching input for subsequent report. However, in daily life, visual search is often guided by non-target objects that are neither externally cued nor reported. For instance, when looking for a pen, we may direct our attention to the office desk where we expect the pen to be. These "anchor objects" (e.g., the desk) thereby guide search for associated objects (e.g., the pen) in scenes. Here, we used fMRI and eye tracking to test whether preparatory activity during visual search represents the target (the pen), the guiding anchor object (the desk) or both. In an anchor-guided search task, participants (N=34) learned associations between targets and anchors and searched for these targets in scenes. To fully dissociate target from anchor processing, target-anchor associations were reversed across different scene contexts. Participants' first fixations were reliably guided towards the target-associated anchor. Importantly, preparatory fMRI activity patterns in lateral occipital cortex (LOC) represented the target-associated anchor rather than the target. Whole-brain analyses additionally identified a region in the right intraparietal sulcus that represented the anchor. Our results show that preparatory activity in visual cortex represents a self-generated guiding template, supporting visual search in structured daily-life environments.

Significance Statement

Quickly finding relevant objects in complex environments is challenging, yet we are remarkably efficient at it. Theories of attention propose that this is mediated by preparatory biases in visual cortex, such that, for example, looking for a blue pen increases the activity of neurons tuned to blue, elongated objects, favoring the processing of such objects once they appear in view. However, guidance is often based on features that are not the target: when looking for a pen, we first direct our attention to the office desk. Here we show that preparatory activity in visual cortex represents the features that are most useful for guidance (e.g., the desk), rather than the target, clarifying the role of preparatory activity in naturalistic search.

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 favor 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., 2018a; Koehler & Eckstein, 2017a, 2017b; Mack & Eckstein, 2011; Zhou & Geng, 2023). 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., 2018b; 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 supports such context-guided search has not been previously investigated. Recent theories of visual search (Wolfe, 2021; Yu 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 whether preparatory activity reflects non-target features for guidance, we designed a naturalistic anchor-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 familiarized 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). After familiarization, 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 reflects the visual features that guide search rather than the features that define the target.

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 familiarized 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. 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 behavioral analyses). 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 center 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.

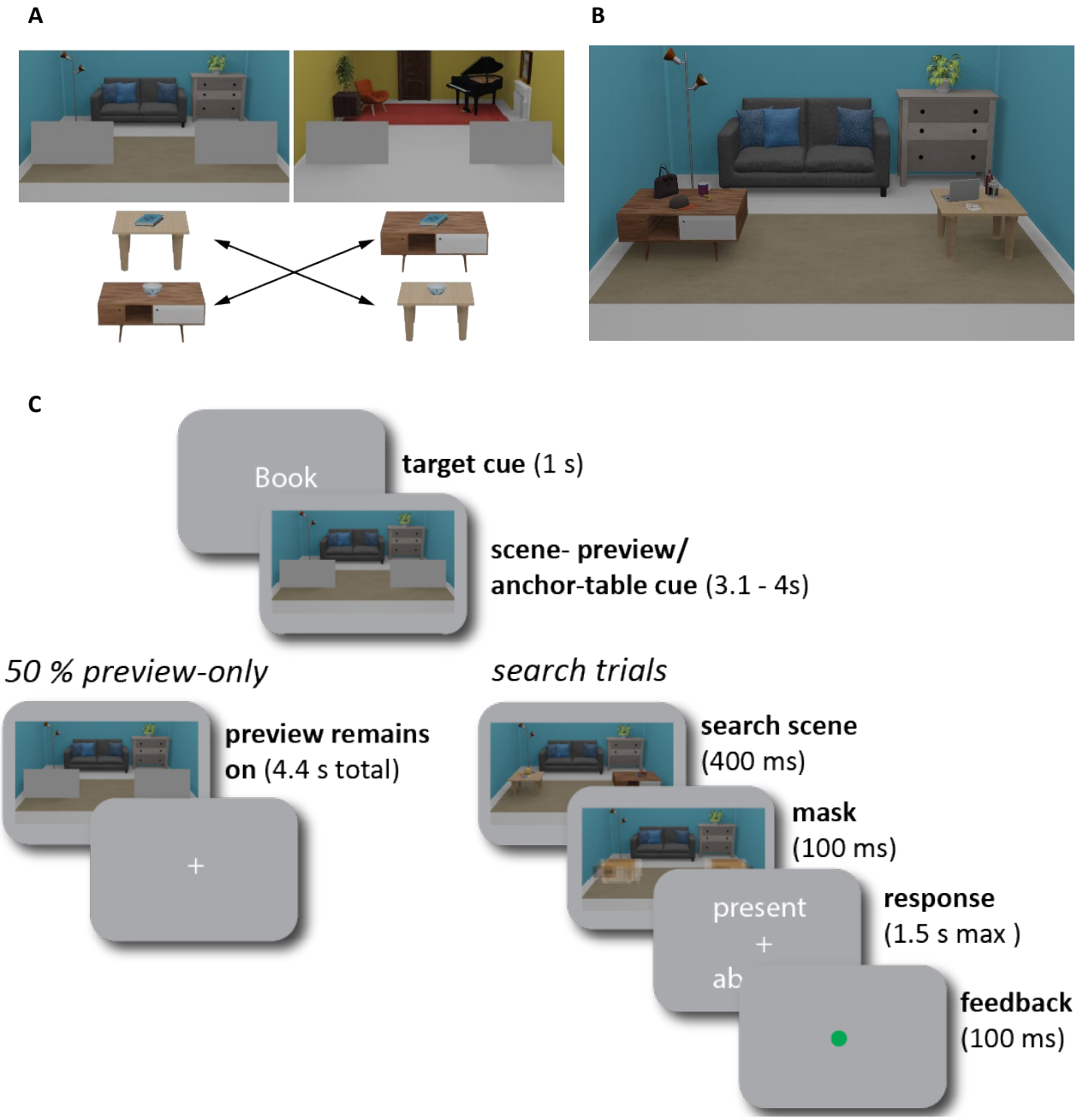

Figure 1: Experimental design and timeline. A Participants learned room-dependent associations between target objects (books and bowls) and two different tables serving as anchor objects. B Example of a search scene. Both tables could either appear left or right in the scene, with multiple objects placed 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.

Eyetracking analysis

First fixations are guided by anchor objects

To test whether anchor objects guided attention, we analyzed 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], $p < 0.001$; Figure 2A,C). 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], $p = 0.03$). This indicates that anchor features indeed guided attention, with an 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' ($r = 0.55$, $p < 0.001$; Figure 2B). This correlation was also significant for target present trials ($r= 0.60$, $p < 0.001$; Figure 2B).

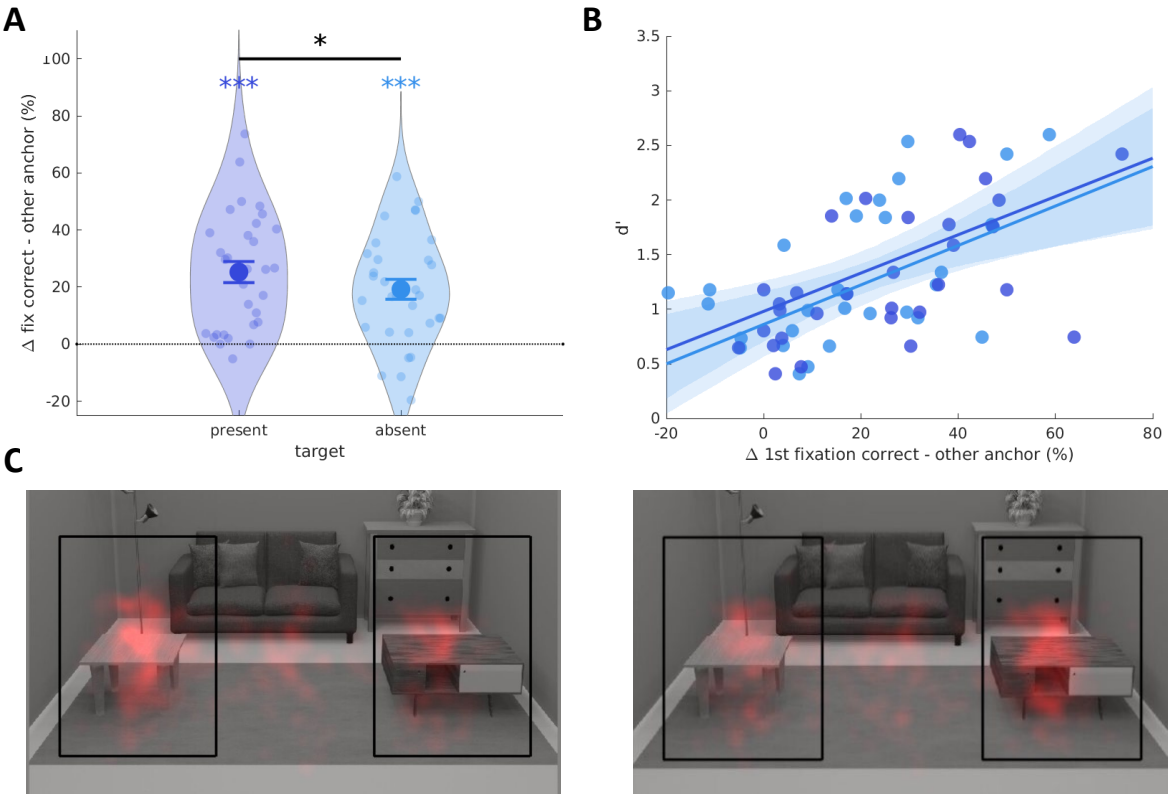

Figure 2: Eye tracking results. **A** *Difference in percentage of first fixations directed towards the correct (relevant) anchor and the other (irrelevant) anchor for target present trials (left) and target absent trials (right). All error bars are SEM. ***, p<.001* **B** *Correlations of d' during 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.*

fMRI results

Preparatory activity in LOC does not reflect the target

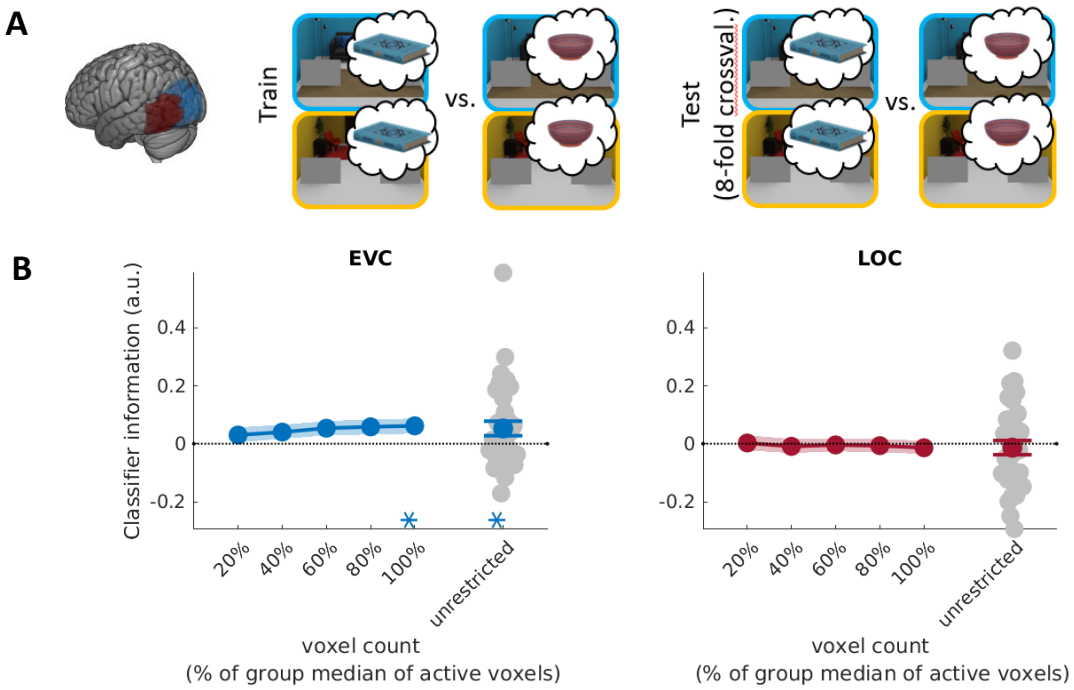

Figure 3: Decoding of the target object within search task runs. ***A*** *Overview of the decoding scheme, decoding the target (book/bowl) from the preview only trials (leave one run out cross-validation). Brain shows anatomical location of the ROIs.* **C** *decoding results in EVC and LOC for all selective 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).*

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. To ensure the robustness of our results to different voxel selections, we repeated this analyses across ROIs of different sizes (see ROI definition). Within EVC, the current search target could be decoded above chance (p = 0.02, CI = [0.0078, 0.11];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 (p = 0.58, CI = [-0.06, 0.04]). Taken together, these results indicate that the search target was not strongly represented in preparatory activity during context-guided search, perhaps because attention (and eye movements) in this task was primarily guided by anchor objects.

Preparatory activity for associated anchor objects

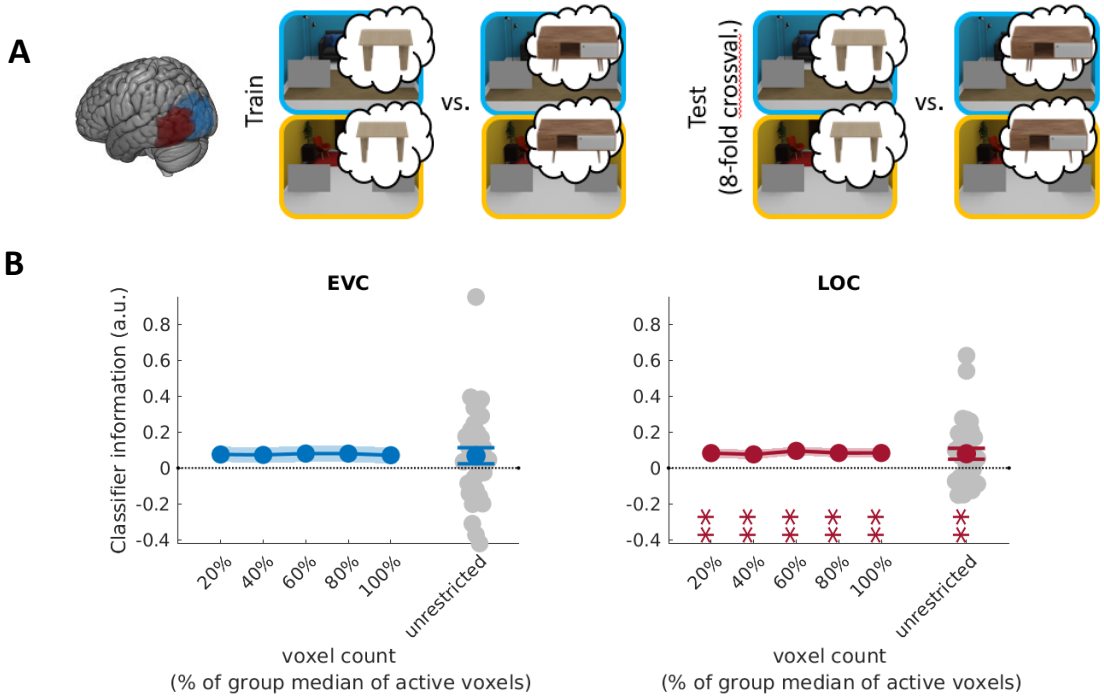

*Figure 4: Decoding of the associated anchor within search task runs **A** Overview of the decoding scheme, decoding the associated anchor (Table 1/Table 2) from the preview only trials (leave one run out cross-validation). Brain shows anatomical location of the ROIs.* **B** *decoding results in EVC 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<0.01; ***, p<.001. For sub-ROIs, asterisks show significance after threshold-free cluster enhancement (TFCE).*

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 ($p = 0.24$, CI = [-0.03, 0.14]; Figure 4B). Importantly, however, the associated anchor could be reliably decoded from LOC ($p = 0.001$, CI = [0.03, 0.15];Figure 4B). Decoding was consistent across all ROI sizes (Figure 4B). 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 learned target-anchor associations.

Cross-scene decoding analysis: Anchor templates in LOC generalize 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).

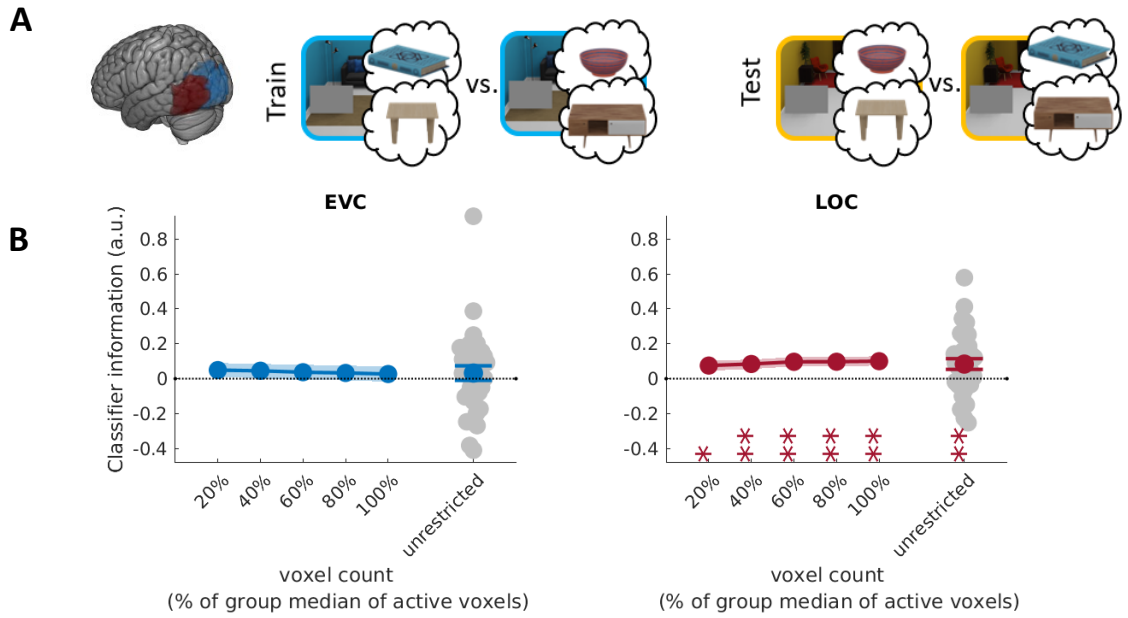

*Figure 5: Decoding of the target-associated anchors across rooms **A** Overview of the decoding scheme, decoding the associated anchor (table 1/table 2) from the preview only trials in one room/scene context, testing on the other (cross-scene decoding). Brain shows the anatomical location of the ROIs.* **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<0.05; **, p<0.01. 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, generalizing 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 generalize across rooms (p = 0.28, CI = [-0.031, 0.13]; Figure 5B). Importantly however, the anchor template in LOC was independent of scene context and the associated target (p = 0.004; CI = [0.03 0.15];Figure 5B). Decoding was consistent across ROI sizes (Figure 5B). 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.

Searchlight analyses

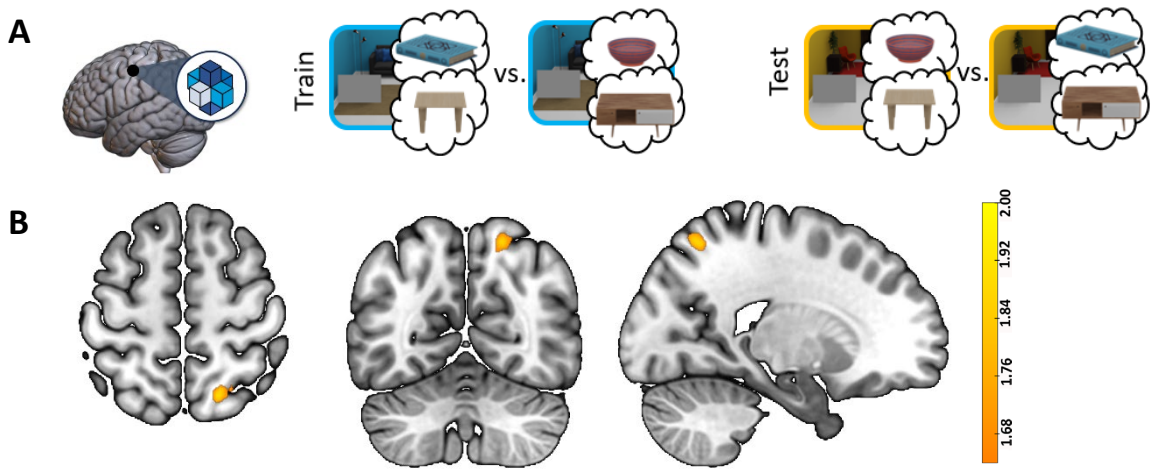

Figure 6: Searchlight for decoding the target-associated anchor across scene-contexts revealed a cluster in the right intraparietal sulcus (x, y, z = 20.1, -63.6, 58, 203 mm³). Color indicates z-scores after threshold free cluster enhancement (TFCE).

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. There were no significant clusters showing negative cross-scene decoding (which would be indicative of preparatory target activity).

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; Yu et al., 2023), and can even be based on entirely different non-target objects, as in the case of guidance by anchor objects. Since those objects are easy to find and predictive of target location, they can help to spatially constrain visual search. Here, we tested the neural basis of such anchor-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), supports such searches by encoding the 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; Yu et al., 2023) that postulate 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 learned associations. Overall, our results show that preparatory activity in visual cortex reflects an internally generated guiding template, independent of external cues or target report.

Our results indicate that preparatory activity in visual cortex can be highly flexible and context-dependent, representing a different anchor object 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 priming or general semantic associations, but rather context-specific associations for guidance. Despite this reversal across contexts, associations were quickly learned, 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 we often expect objects to 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 learned 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 (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; Summerfield et al., 2006). Similar mechanisms may be involved when recently learned 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 learned 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 et al., 2019).

The searchlight analysis revealed that anchor-specific preparatory activity was also present in the right intraparietal sulcus. Similar to LOC, this activity generalized 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 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). 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 table 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.

In conclusion, we show that visual search in naturalistic scenes 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.

Method

Behavioral, eyetracking and fMRI data, stimuli and experiment/analysis code will be available on the Radboud Data Repository following publication.

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: 5 due to at-chance accuracy in the search task (determined by a 1-sided binomial test with $\alpha = 0.05$), 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 (8 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.

Localizer task

To localize voxels that were selective to the targets and/or anchors, we also included four runs of a separate localizer 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 centers in the search task) or a single object in the center of fixation (as they would appear after an eye movement to fixate them). Each run consisted of 32 miniblocks (4 repetitions of 8 conditions: book/bowl/Table 1/Table 2 x central/peripheral location).

Within each miniblock, participants saw 8 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, memorized 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 localizer task was presented interleaved with the search task, with one localizer followed by two search task runs. For the first three subjects, a different localizer 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 generalization between visually evoked responses in this localizer 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 memorized size in this localizer 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 learned 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 center, and had to indicate the location of the associated table given target and scene-context using a keyboard. Participants received feedback after each trial and were reminded of the associations after the task if needed.

Stimuli

640 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 color 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). 320 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 localizer, 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 center at the same height. For the target objects, 20 images per localizer condition showed all category exemplars used in the search task. For the anchor objects, 10 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 9-point calibration procedure (resorting to a 5 point calibration in case no accurate 9-point calibration was possible) and recalibrated if required.

Eye tracking AOIs

Three rectangular areas of interest (AOIs) were defined for analyzing fixations during the search scene. First, a center AOI around the fixation cross (4.5 x 10.5°) to ensure fixation was in the image center before the search scene appeared and one centered 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 center 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 localizer 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 (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 preprocessed images. For the search task, the 4 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 localizer 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 . We assumed that any preparatory bias would be strongest in those voxels responding strongly to visual features of either the target or anchor objects, and therefore included specifically those voxels for our analysis. Within each ROI, we selected voxels that discriminated between the targets or anchor objects, when they were visually presented during the independent localizer 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 ($p < 0.05$, 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-selective voxels within one ROI-mask, and results are based on these

combined ROI maps. On average, these masks included 1289.85 voxels in EVC (sd 58.17) and 917.32 voxels (sd 50.72) in LOC within each hemisphere.

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 Supplementary Material Supplementary Materials 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 Supplementary Materials Figure S2 for decoding results within individual hemispheres).

Classification was either performed in (1) a leave-one-run out 8-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.

Classifier performance was calculated based on the distance from the decision boundary, providing a more sensitive and continuous estimate for each beta compared to a binary label (Walther et al., 2016), and 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).

Searchlight analysis

To complement the previous ROI analyses, we ran an additional searchlight analysis, testing for generalization of the anchor template across scenes (1-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 (2002)) to decode the targets or anchors from the localizer betas. Voxels with significant decoding for either the target or anchor at the group-level (alpha = 0.05) were then used to create a mask for the cross-decoding searchlight. For all searchlights, spheres had a radius of 5 voxels (resulting in around 524 voxels in total). .

Statistical tests

To test for differences between conditions in the behavioral/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 2-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 neighboring

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.

Acknowledgements

We would like to thank Cara Struckmann for help with stimulus creation, Giacomo Aldegheri for help with Blender programming, and Surya Gayet for helpful comments on this manuscript. This project received funding from 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.

Author Contributions

M.L, F. L. and M.P. designed research, M.L performed research and analyzed data. M.L. F.L. and M.P. interpreted the data. M.L, F.L. and M.P. wrote the paper.

References

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/https://doi.org/10.1111/nyas.13320

Bichot, N. P., Rossi, A. F., & Desimone, R. (2005). Parallel and serial neural mechanisms for visual search in macaque area V4. *Science*, *308*(5721), 529–534. https://doi.org/10.1126/science.1109676

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Võ, M. L. H. (2018a). 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., Draschkow, D., Dienhart, E., & Võ, M. L. H. (2018b). 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

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., 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

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

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

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

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

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

Favila, S. E., & Aly, M. (2023). *Hippocampal differentiation of competing memories predicts the precision of preparatory coding in visual cortex*.

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

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

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., 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

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

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

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. https://www.biorxiv.org/content/10.1101/2023.09.11.557257v1%0Ahttps://www.biorxiv.org/content/10.1101/2023.09.11.557257v1.abstract

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

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*, 1–12. https://doi.org/10.1080/13506285.2021.1915902

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

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

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

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

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

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

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

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

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

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

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

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 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

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

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

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

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

Yu, X., Zhou, Z., Becker, S. I., Boettcher, S. E. P., & Geng, J. J. (2023). Good-enough attentional guidance. *Trends in Cognitive Sciences*, *xx*(xx), 1–13. https://doi.org/10.1016/j.tics.2023.01.007

Zhou, Z., & Geng, J. J. (2023). Learned associations serve as target proxies during difficult but not easy visual search. *PsyArxiv*. https://doi.org/10.31234/osf.io/y26qx

Supplementary Materials

Behavioral analyses

Overall behavioral accuracy in the search task was 72.46% (sd 8.91). 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: $F(1,33) = 1.04$, $p = 0.32$, $\eta_p^2 = 0.03$; no main effect of Scene context: $F(1,33) = 0.07$, $p = 0.79$, $\eta_p^2 < 0.01$; no Target x Scene context interaction: $F(1,33) = 3.06$, $p = 0.07$, $\eta_p^2 = 0.09$). 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 ($F(1,33) = 5.84$, $p = 0.02$, $\eta_p^2 = 0.15$). While performance for the two tables differed in the yellow living room (CI = [4.86, 10.85] $p < 0.001$) they did not differ for the blue living room (CI = [-3.2, 5.19], $p = 0.58$). Altogether, these findings indicate that there were no consistent accuracy differences across targets, anchors, or rooms.

fMRI decoding for target- and anchor-selective voxels

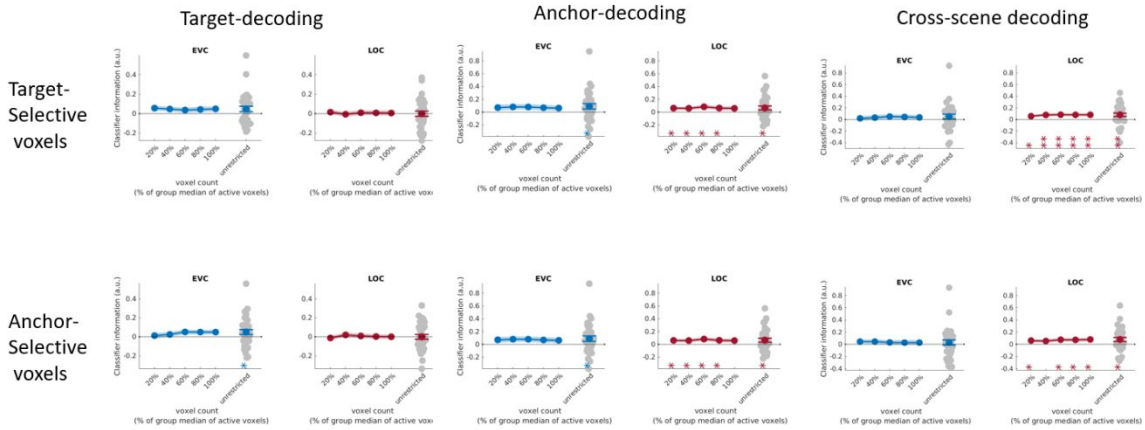

*Figure S1: Decoding results shown separately for target and anchor-selective voxels in EVC and LOC, for all selective voxels of individual participants (unrestricted-ROI) and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (unrestricted) ROI. All error bars are SEM. *, p<0.05; **, p<0.01*

fMRI decoding for individual hemispheres

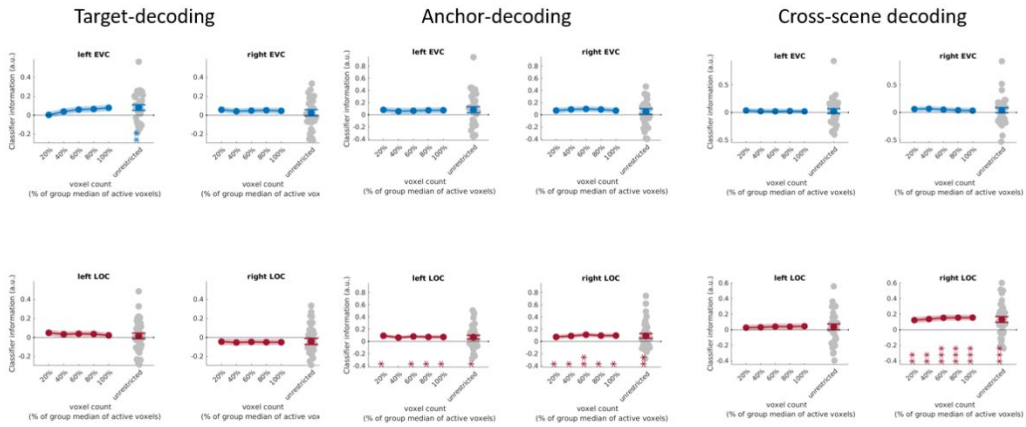

*Figure S2: Decoding results shown separately for left and right hemisphere in EVC and LOC, all selective voxels of individual participants (unrestricted-ROI), and sub-ROIs of different sizes. Grey dots show decoding for individual participants in the largest (all) ROI. All error bars are SEM. *, p<0.05; **, p<0.01*