NeuroImage 272 (2023) 120053

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/neuroimage

Independent spatiotemporal effects of spatial attention and background clutter on human object location representations

Monika Graumann[a,b,*], Lara A. Wallenwein[c], Radoslaw M. Cichy[a,b,d]

[a] Department of Education and Psychology, Freie Universität Berlin, 14195 Berlin, Germany
[b] Berlin School of Mind and Brain, Faculty of Philosophy, Humboldt-Universität zu Berlin, 10117 Berlin, Germany
[c] Department of Psychology, Universität Konstanz, 78457 Konstanz, Germany
[d] Bernstein Center for Computational Neuroscience Berlin, 10115 Berlin, Germany

ARTICLE INFO

Keywords:
fMRI
EEG
Attention
Decoding
Multivariate pattern analysis
Object recognition

ABSTRACT

Spatial attention helps us to efficiently localize objects in cluttered environments. However, the processing stage at which spatial attention modulates object location representations remains unclear. Here we investigated this question identifying processing stages in time and space in an EEG and fMRI experiment respectively. As both object location representations and attentional effects have been shown to depend on the background on which objects appear, we included object background as an experimental factor. During the experiments, human participants viewed images of objects appearing in different locations on blank or cluttered backgrounds while either performing a task on fixation or on the periphery to direct their covert spatial attention away or towards the objects. We used multivariate classification to assess object location information. Consistent across the EEG and fMRI experiment, we show that spatial attention modulated location representations during late processing stages (>150 ms, in middle and high ventral visual stream areas) independent of background condition. Our results clarify the processing stage at which attention modulates object location representations in the ventral visual stream and show that attentional modulation is a cognitive process separate from recurrent processes related to the processing of objects on cluttered backgrounds.

1. Introduction

Spatial attention helps us to focus visual processing on the relevant portions of the visual field while ignoring its irrelevant portions (Desimone and Duncan, 1995). For example, spatial attention helps during navigation to determine where in visual space objects are located, allowing us to avoid obstacles and to reach desired targets better.

In spite of ardent research in humans and other primates over more than 50 years (Desimone and Duncan, 1995; Mangun, 1995; Hillyard et al., 1998a, 1998b; Luck et al., 2000; Carrasco, 2011; Wolfe et al., 2011; Squire et al., 2013; Maunsell, 2015), no unified view has pinpointed attentional modulation of object location to a specific processing stage in the visual hierarchy. Previous research yielded contradictory results. Considering the temporal emergence of attentional effects, some studies which cued the stimulus location found attentional modulation early (Van Voorhis and Hillyard, 1977; Mangun, 1995; Hillyard et al., 1998a, 1998b; Hopfinger et al., 2000; Luck et al., 2000) within a time window that corresponds to the initial bottom-up response within the first 150 ms (Lamme and Roelfsema, 2000; VanRullen and Thorpe, 2001; Fahrenfort et al., 2007; Camprodon et al., 2010; Koivisto et al., 2011). In contrast, studies with no pre-stimulus location cue found such effects predictably later (Wyatte et al., 2014; Groen et al., 2016; Kaiser et al., 2016a; Battistoni et al., 2020). Similarly, considering the locus in the visual processing hierarchy some studies found attentional modulation already in V1 (Roelfsema et al., 1998; Martínez et al., 2001; Noesselt et al., 2002; Khayat et al., 2006; Lakatos et al., 2008; Briggs et al., 2013; Herrero et al., 2013) while other studies found such effects only or predominantly in higher-level brain regions (Buffalo et al., 2010; Peelen and Kastner, 2011; Kay et al., 2015).

The contradiction might be resolved when considering together the processing stage at which object location representations emerge, the object's viewing conditions and the timing of attentional engagement. For example, studies detecting attentional modulation at early processing stages often used low-level stimuli on blank backgrounds such as Gabor patches (Hillyard et al., 1998b, 1998a; Martínez et al., 2001; Noesselt et al., 2002; Briggs et al., 2013). In contrast, studies finding attentional modulation at later processing stages used realistic, high-level stimuli such as objects and scenes (Peelen and Kastner, 2011; Kay et al., 2015; Kaiser et al., 2016b; Battistoni et al., 2020). However, the general pattern of results suggests that even with low-level

* Corresponding author.
E-mail address: monika.graumann@fu-berlin.de (M. Graumann).

https://doi.org/10.1016/j.neuroimage.2023.120053.
Received 21 September 2022; Received in revised form 21 March 2023; Accepted 23 March 2023
Available online 24 March 2023.

stimulation, attentional modulation is lower in V1 than in higher areas like V4 (Tootell et al., 1998; Kastner et al., 1999; Buffalo et al., 2010; Carrasco, 2011). Thus, previous studies varied both in pre-stimulus location cueing and viewing conditions. Cueing might influence the timing of the measured onset of attentional modulation because attention can be engaged before stimulus onset and therefore results in different latencies of attentional modulation onset. Viewing conditions might influence the results because recent research has shown that they influence the processing stage at which object location representations emerge. For example, object location representations emerge early for objects on blank and late on cluttered backgrounds (Hong et al., 2016; Graumann et al., 2022). These findings are important for two reasons: First, they show that with no clutter, object location likely reflects simple retinotopic mapping, while with clutter, a more complex processing cascade is necessary to encode an object's location. This stands in contrast to traditional theories of object perception (Ungerleider and Haxby, 1994; Milner and Goodale, 2006). Second, the surroundings of an object modulates the employment of spatial attention: spatial attention is more relevant for the localization of objects in clutter than in isolation (Treisman and Gelade, 1980; Wolfe, 1994).

Here we set out to untangle the complex link between the processing stage at which object location representations emerge, its viewing conditions, and the processing stage of attentional modulation when the location is not cued in advance.

Our hypotheses are as follows. We set the stage by hypothesizing based on recent findings that the processing stage at which object location representations emerge depends on the object's viewing conditions in particular its background (Graumann et al., 2022) independent of spatial attention. This replication hypothesis was termed $H_{Replication}$ (abbreviated H_R; Fig. 1A,C).

On this basis we then theorize how an objects background impacts when (in time with respect to stimulus onset) and where (in the cortical processing hierarchy) attention modulates location representations. We propose two alternative hypotheses.

The first hypothesis is that attention and background interact: attention dynamically modulates location representations at the processing stage at which they first emerge, resulting in an interaction between background and attention ($H_{Dynamic}$, abbreviated H_D; Fig. 1B,D). The alternative hypothesis is that attention modulates location representations statically and always during a late processing stage (Wyatte et al.,

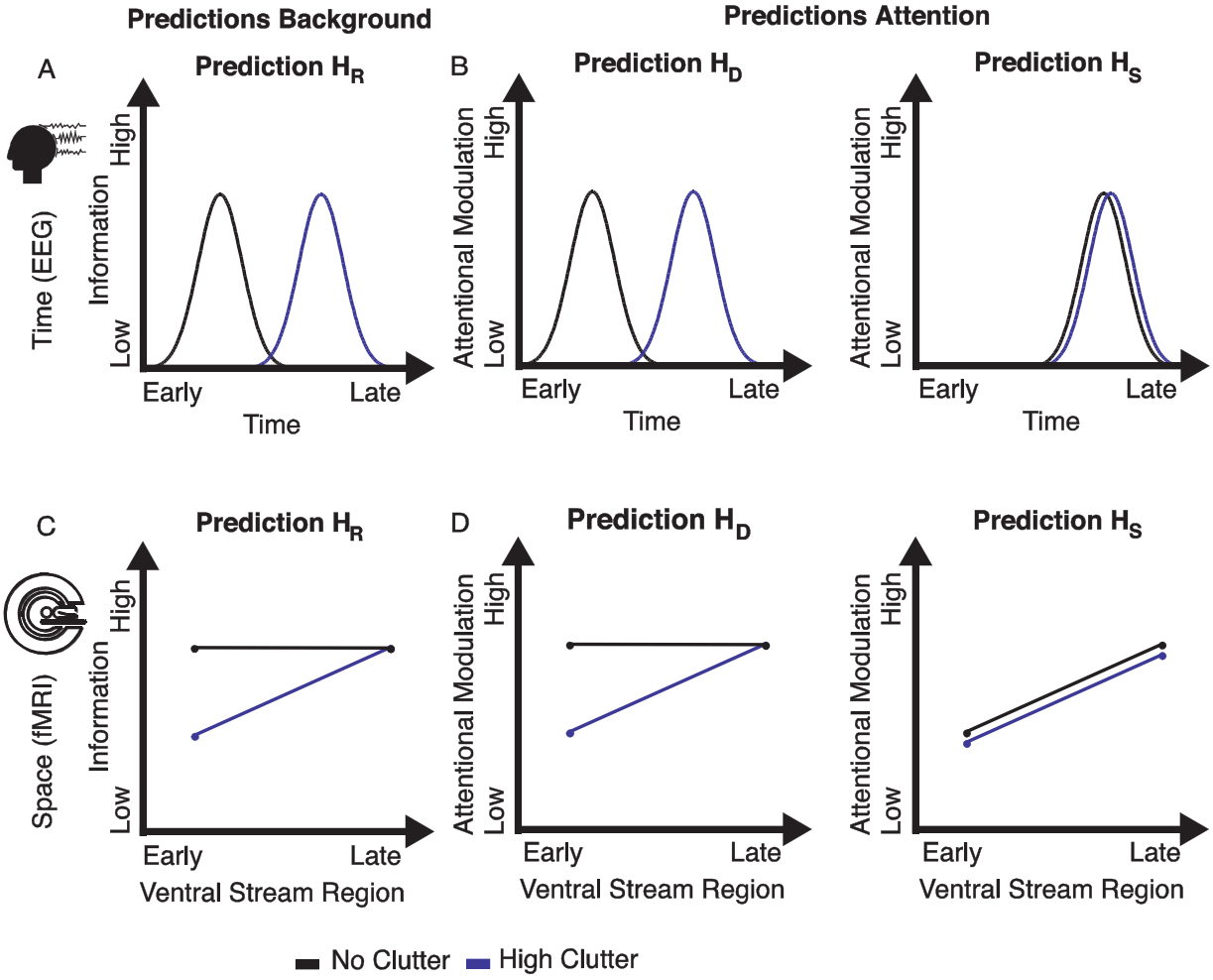

Fig. 1. Experimental predictions based on hypotheses. **A,** Predictions for the effect of background on location information in the EEG experiment. H_R predicts a delay in time for location information with high clutter compared to no clutter. **B,** Predictions for the effect of attention on location information in the EEG experiment. Predictions are based on H_R in A. H_D predicts that the time point when attentional modulation is highest depends on background: attentional modulation is highest at time points when location information is highest, depending on background condition. H_S predicts that attentional modulation is always highest during late processing stages, independent of background condition. **C,** Predictions for the effect of background on location information in the fMRI experiment. H_R predicts an increase along the ventral stream for location information with high clutter compared to no clutter. **D**, Predictions for the effect of attention on location information in the fMRI experiment. Predictions are based on H_R in C. H_D predicts that the region where attentional modulation is highest depends on background: attentional modulation is highest in regions where location information is highest, depending on background condition. H_S predicts that attentional modulation always increases along the ventral stream, independent of background condition.

2014; Kay et al., 2015; Groen et al., 2016; Kaiser et al., 2016a; Battistoni et al., 2020), independent of the background (H_{Static}, abbreviated H_S; Fig. 1B,D).

We investigated these hypotheses in an integrated research project consisting of an EEG and fMRI experiment in combination with multivariate pattern analysis methods. We manipulated background by presenting objects on backgrounds of different clutter levels, and attention by task instruction that attracted or diverted spatial attention from an object's location. Here we defined location representations as the neural response patterns of an area to an object in a given retinotopic location, thereby linking brain activity to visual object localization (Kriegeskorte and Diedrichsen, 2019). Thus, we focused on retinotopic rather than spatiotopic location representations.

To anticipate, we first confirmed H_R, i.e., object location representations emerged later in time and space when the object appeared on a cluttered background than on a blank background, independent of attention. We then found strong empirical support for H_S. That is, attention modulates object location representations late in both time and space, independent of background.

2. Materials and methods

2.1. Participants in EEG and fMRI experiment

27 participants completed the EEG experiment. One participant was excluded because of technical problems, resulting in 26 participants (mean age 26.42 years, *SD*=4.12, 19 female) included in the final EEG study. 23 participants completed the fMRI experiment, out of which one also participated in the EEG experiment. Three participants were excluded because they did not complete the whole experiment, resulting in 20 participants (mean age 26.71 years, *SD*=4.48, 13 female) included in the final fMRI study. Sample sizes were chosen to be comparable to a previous similar study (Graumann et al., 2022).

All participants had no history of neurological disorders and normal or corrected-to-normal vision. Participants provided written informed consent prior to the studies and participation was compensated with payment or course credit. The study was conducted in accordance with the Declaration of Helsinki and the ethics committee of the Department of Education and Psychology of the Freie Universität Berlin approved the study in advance.

2.2. Experimental design

2.2.1. EEG experimental design

The experimental design in the EEG study comprised the four factors object category (animals, cars, faces, chairs, Fig. 2A left, with 3 exemplars per category), location (left up, left bottom, right bottom, right up, Fig. 2A left center), background (no and high clutter, Fig. 2A center) and attention (on periphery or on fixation, Fig. 2A right center). These four factors were fully crossed, to investigate them independently of each other. Specifically, including the factor category served to systematically analyze location information that was independent of potentially confounding category information in a cross-classification approach (see section 2.7). In total, this created 192 individual conditions (12 object exemplars × 4 locations × 2 background conditions × 2 attention conditions). For further analysis, data was collapsed across exemplars, so that data was analyzed at the level of category. Thus, the number of conditions for further analysis was 64 (4 categories × 4 locations × 2 background conditions × 2 attention conditions, Fig. 2A right).

2.2.2. fMRI experimental design

The experimental factors in the fMRI experiment were the same as in the EEG experiment, but there were two instead of four levels for the factors category (cars, faces) and location (left, right; Fig. 2B). This resulted in 48 individual conditions (6 object exemplars × 2 locations × 2 background conditions × 2 attention conditions). As in the EEG experiment, the inclusion of the factor category served to systematically analyze location information that was independent of potentially confounding category information in a cross-classification approach (see section 3.7). For further analysis, data was likewise collapsed across exemplars, so that data was analyzed at the level of category. Thus, the number of conditions for further analysis was 16 (2 categories × 2 locations × 2 background conditions × 2 attention conditions).

2.3. Stimulus set generation

2.3.1. Stimulus set generation: EEG experiment

The experimental design in the EEG study comprised 96 individual stimulus conditions shown in each attention condition, as detailed in the previous section. To create these stimuli, each exemplar was superimposed onto backgrounds with or without scene images in four locations. First, to position object exemplars onto the four image locations, we projected the 3D rendered objects onto to the four quadrants of the screen (Fig. 2A, left center). Rendered objects did not extend beyond a quadrant. Each object's center was positioned 3° from the vertical and 3° from the horizontal central midline (i.e., 4.2° diagonally from image center to fixation, Fig. 2A right), subtending 2.4° (*SD*=0.4) in vertical and 2.2° (*SD*=0.6) in horizontal extent.

Second, each exemplar in each location was superimposed onto a background with no and with high clutter (Fig. 2A, center; the backgrounds shown here are comparable to the original backgrounds used in the experiments). We chose the background conditions no and high clutter to compare visual stimuli with low and high image complexity, respectively (Groen et al., 2018). The no clutter condition was a uniform gray background. In the high clutter condition, we selected 60 natural scene images from the Places365 database (http://places2.csail.mit.edu/download.html) that did not contain objects of the categories included in our experimental design (i.e., no animals, cars, faces, chairs) and were highly cluttered (as defined by 10 independent subject ratings; for methods and results see Graumann et al., 2022). We converted the images to grayscale and superimposed a circular aperture of 15°. Original backgrounds are not shown because of copyright reasons but are available here: https://osf.io/85sak/?view_only=db183dde8f4b406aaba5dfc0dd0ae67d.

From the set of 60 scene images, we selected 48 scene images to go with the 48 stimulus conditions within the high clutter condition (12 exemplars × 4 locations). To avoid systematic congruencies between objects and background images within the high clutter condition, stimulus conditions and backgrounds were randomly paired for each of the 20 runs into which the EEG experiment was divided (see below). Together with the 48 stimulus conditions in the no clutter condition, this resulted in 96 individual images per run. The 12 remaining scene images from the set of 60 were used to create catch trials. Images were not normalized for overall luminance and luminance was higher in the no (109) than in the high clutter condition (100). The average contrast across images was higher with high clutter (59) than with no clutter (36).

2.3.2. Stimulus set generation: fMRI experiment

Stimulus set generation for the fMRI experiment was equivalent to the EEG experiment, with the difference that objects were positioned on two instead of four image locations (Fig. 2B) 4.2° to the left or right of the image's center. In the fMRI experiment, each background condition had 12 individual stimulus conditions (6 exemplars × 2 locations). In combination with the 12 stimulus conditions in the no clutter condition, this resulted in 24 individual images per run. The remaining scene images from the set of 60 were used to create 24 catch trials (1 catch object × 12 scene images × 2 locations), which were randomly presented during the fMRI experiment. Images were not normalized for overall luminance and luminance was higher in the no (109) than in the high clutter condition (100). The average contrast across images was higher with high clutter (59) than with no clutter (36).

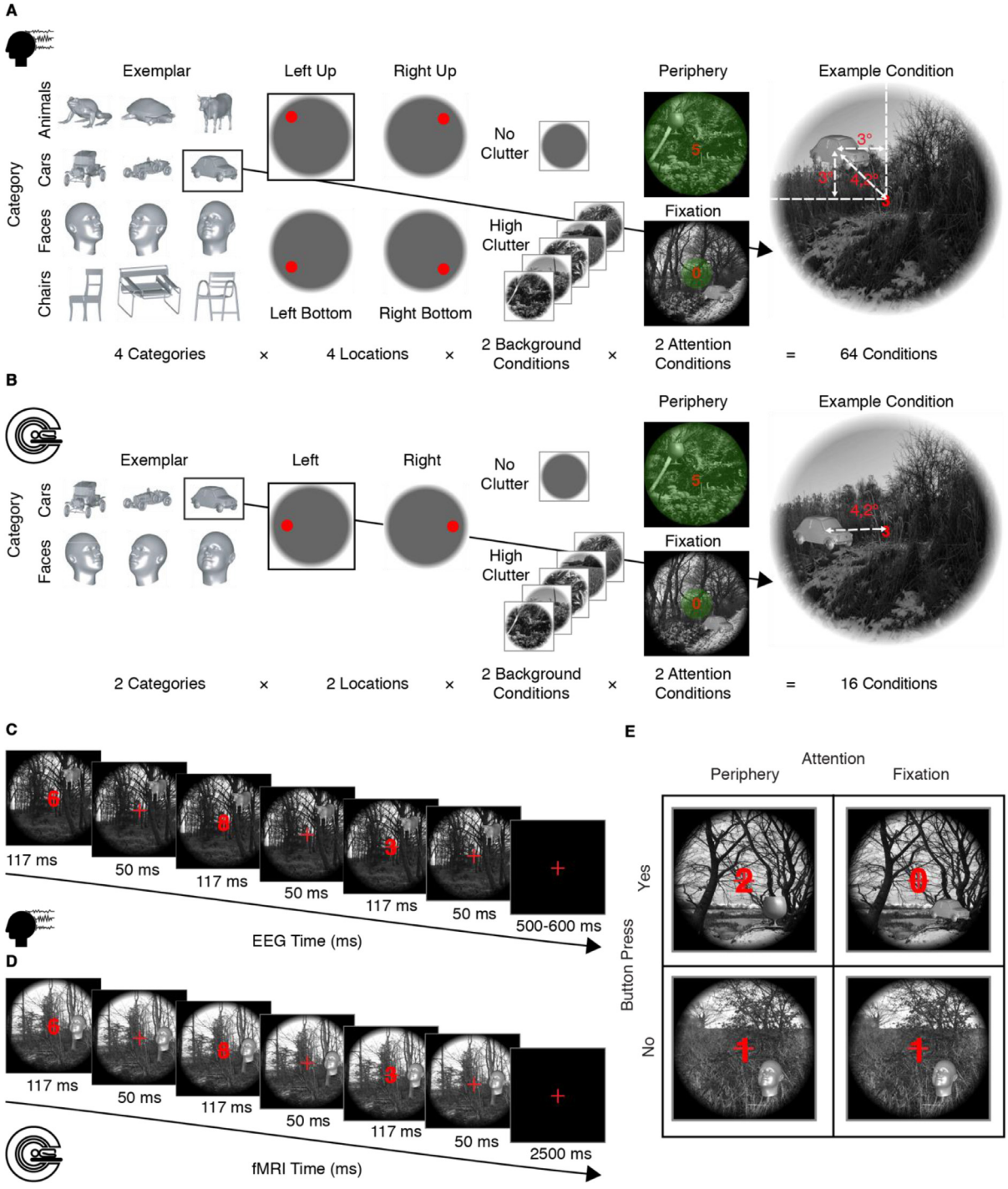

Fig. 2. Experimental design and tasks. **A,** Experimental design in the EEG experiment. We used a fully crossed design with factors: object category, location, background and attention. Green translucent circles represent attentional width. **B**, Experimental design in fMRI experiment. The design was equivalent to the EEG experiment, except that the factors category and location had two levels. **C,** Trial timing and example condition in EEG experiment. **D,** Trial timing and example condition in fMRI experiment. **E,** Tasks. In the peripheral attention condition (left) participants responded with button press when a glass appeared in the periphery, while fixating their gaze on the central cross. Digits presented on fixation were task-irrelevant. In the fixation attention condition (right) participants responded with button press when the digit 0 appeared on fixation, while fixating on the central cross. Objects in the periphery were irrelevant in this task. Visual stimulation was the same in both tasks on regular trials (see bottom row 'Button press: no').

2.4. Experimental procedures

2.4.1. EEG main experiment

Each of the 26 participants completed one EEG recording session with 20 runs (run duration: 277 s). Overall, the EEG session lasted for 92 min. Participants performed attention tasks on separate runs. The EEG recording session consisted of 10 periphery attention runs and 10 fixation attention runs in randomized order. Within each attention condition, there were 96 individual stimulus conditions (12 exemplars × 4 locations × background conditions). Runs consisted of the presentation of regular trials and catch trials. In each run, there were 192 regular trials, representing the 96 stimulus condition images presented twice.

These trials formed the basis for further analysis. On regular trials, digits between 1 and 9 were overlaid for 117 ms each, followed by a 50 ms presentation of the image and fixation cross after each digit (Fig. 2C). In total, stimuli were presented for 0.5 s followed by 0.5 or 0.6 s of ISI (equally probable; Fig. 2C). Participants were asked to fixate their eyes on the central cross at all times.

On catch trials, a target was presented to which participants were asked to respond with button press (Fig. 2E). These trials were excluded from the analyses. Catch trials were presented on every 3rd to 5th trial (equally probable, in total 48 per run). Participants were instructed to respond with button press to catch trials and to blink their eyes to minimize eye blink contamination on subsequent trials. The ISI was 1 s on catch trials to avoid contamination of movement and eye blink artefacts on subsequent trials.

In the periphery and the fixation attention condition different trials were task-relevant catch trials. In the periphery attention condition, catch trials were trials during which a target object (a glass) was presented (Fig. 2E). The target could be presented at any of the four locations and on any type of background. Digits on fixation were task-irrelevant in this attention condition. In the fixation attention condition, catch trials were trials during which the digit 0 appeared among any of the 3 digits that were presented on fixation during a single trial (Fig. 2E). The presented object in the periphery was task-irrelevant in this attention condition (Fig. 2E). The digit 0 never appeared on periphery attention runs and the glass never appeared on fixation attention runs.

2.4.2. fMRI main experiment

Each of the 20 participants completed one fMRI recording session with 20 runs (run duration: 288 s). Overall, an fMRI recording in the main experiment lasted for 96 min. Each of the 24 images of the stimulus set was shown 3 times in random order without back-to-back repetitions in each run. On each trial, the image was presented for 0.5 s at the center of a black screen. The inter-stimulus-interval (ISI) was 2.5 s (Fig. 2D). Images were overlaid with a red central cross for fixation. Participants were instructed to fixate their eyes on this cross throughout the experiment. Every 3rd to 5th trial (equally probable, in total 18 per run) a catch trial was presented. The tasks in the attention conditions and the catch objects were identical to the EEG experiment (Fig. 2E). Catch trials were excluded from further analysis.

2.4.3. fMRI localizer experiment

Prior to the main fMRI experiment, participants completed a separate localizer run to define ROIs in early visual, dorsal and ventral visual stream. We presented images from three categories: faces, objects, and scrambled objects. Each image showed identical versions of the same object located left and right of fixation to stimulate the same retinotopic regions of visual cortex as the objects in the main experiment.

The localizer run lasted for 384 s, during which we presented 6 stimulation blocks. Each block was 16 s long with presentations of 20 different objects from one of the three categories (500 ms on, 300 ms off) block-wise. Each block included two one-back image repetitions to which participants had to respond to with a button press. The order of these blocks was first order counterbalanced: triplets of stimulation blocks were presented in random order and interspersed with blank background blocks.

2.5. EEG acquisition and preprocessing

To record EEG data, we used the EASYCAP 64-channel system with a Brainvision actiCHamp amplifier at a sampling rate of 1000 Hz and with an online filter between 0.03 and 100 Hz. The signal was online re-referenced to FCz. Electrode placement followed the standard 10–10 system. Data was preprocessed offline with the EEGLAB toolbox version 14 (Delorme and Makeig, 2004). This comprised a low-pass filter with a 50 Hz cut-off, trial epoching in a peri-stimulus time window between −100 ms and 999 ms, and baseline-correction by subtracting the mean of the 100 ms prestimulus time window from the entire epoch. We used independent component analysis (ICA) to clean the data from ocular and muscular artefacts. To guide the visual inspection of components for removal we used SASICA (Chaumon et al., 2015). To identify horizontal eye movement components, we used external electrodes from the horizontal electrooculogram (HEOG). We detected blink artefact and vertical eye movements using the two frontal electrodes Fp1 and Fp2. On average, we removed 18 (*SD*=5) components per participant. We finally applied multivariate noise normalization on the preprocessed data to improve the signal-to-noise ratio and reliability of the data (Guggenmos et al., 2018).

2.6. Preprocessing and univariate fMRI analysis

2.6.1. fMRI acquisition and preprocessing

MRI data was recorded using a 12-channel head coil on a 3T Siemens Tim Trio Scanner (Siemens, Erlangen, Germany). The structural image was acquired with a T1-weighted sequence (MPRAGE; 1-mm^3 voxel size). To acquire functional data for the main experiment and the localizer run, we ran a T2*-weighted gradient-echo planar sequence (TR=2, TE=30 ms, 70° flip angle, 3-mm^3 voxel size, 37 slices, 20% gap, 192-mm field of view, 64 × 64 matrix size, interleaved acquisition) on the entire brain. fMRI data was preprocessed using SPM8 (https://www._l.ion.ucl.ac.uk/spm/), involving realignment, coregistration and normalization to the structural MNI template brain. We smoothed functional data from the localizer run with an 8 mm FWHM Gaussian kernel, but the data from the main experiment were not smoothed.

2.6.2. Univariate fMRI analysis

We modelled the fMRI responses of the experimental conditions at the level of category. This was done for each run in the main experiment separately using a general linear model (GLM). We entered onsets and durations of stimulus presentations per category, pooling exemplars and repetitions. Thus, each GLM was estimated based on 9 trials (3 exemplars × 3 condition repetitions per run) and was convolved with the hemodynamic response function (hrf). We further entered movement parameters into the GLM as nuisance regressors. This resulted in 8 beta maps per attention condition run (2 categories × 2 locations × 2 backgrounds). For each run, we converted GLM parameter estimates into *t*-values by contrasting each parameter estimate against the implicit baseline for each condition. This resulted for each participant and attention condition run separately in 8 (2 categories × 2 locations × background conditions) *t*-value maps per condition. In sum, this resulted in 8 *t*-value maps per 10 runs, per 2 attention conditions and per participant, which were later used in the classification analysis.

For the fMRI responses to the localizer experiment, we modelled the responses to objects, faces and scrambled objects by entering block onsets and durations as regressors of interest and movement parameters as nuisance regressors into the GLM and convolved them with the hrf. This resulted in three parameter estimates which we used to generate two contrasts that formed part of ROI definitions. The first contrast was defined as objects and scrambled objects > baseline and was used to localize activations in early, mid-level ventral and dorsal visual regions (V1, V2, V3, V4, IPS0, IPS1, IPS2, SPL). The second contrast was defined as objects and faces > scrambled objects and was used to localize activations in object-selective area LOC. Overall, this yielded two *t*-value maps for the localizer run for each participant.

2.6.3. Definition of regions of interest

To define ROIs, we first applied anatomical masks and then selected voxels using appropriate contrasts from the functional localizer run. In detail, we first defined ROIs using anatomical masks from a probabilistic atlas (Wang et al., 2015) and combined these for both hemispheres. We

included three masks in early visual cortex V1, V2 and V3. V4 and LOC served as ROIs in mid- and high-level ventral visual cortex. We also included four ROIs from dorsal visual cortex: IPS0, IPS1, IPS2 and SPL. We removed all overlapping voxels from these masks to avoid overlap between ROIs. The second step entailed selecting the most activated voxels of the participant-specific *t*-value maps of the localizer run within the previously defined anatomical masks. To keep the number of voxels constant between ROIs and participants and improve comparability across ROIs, we determined a fixed voxel number across ROIs and participants instead of using a threshold to avoid a variable number of voxels and thus power. For this, we determined the smallest ROI in any participant when overlaying the localizer *t*-value maps and the anatomical masks. This resulted in a minimum ROI size of 288 voxels. This was then the fixed number of highest activated voxels to select of the participant-specific localizer *t*-value maps within all anatomical masks and participants. To select voxels in LOC we used the objects > scrambled contrast and to select voxels in the remaining ROIs we used the objects & scrambled objects > baseline contrast. This resulted in ROI definitions that were specific to each participant with an equal number of voxels across ROIs and participants.

2.7. Object location classification from brain measurements

To measure location information in time using EEG and in space using fMRI, we applied multivariate classification (Carlson et al., 2011a; Cichy et al., 2011, 2013; Isik et al., 2014) of object location. Since object location and object category have partly overlapping neural fingerprints in time and space (Cichy et al., 2011; Graumann et al., 2022), we applied a cross-classification scheme that avoided location information results to be confounded with category information (Carlson et al., 2011b; Isik et al., 2014). For this, we cross-classified locations across categories, meaning that during each classification of a given location pair, we trained and tested on different object categories. For all classification analyses described, we employed a binary c-support vector classification (C-SVC) with a linear kernel from the libsvm toolbox (Chang and Lin, 2011) (https://www.csie.ntu.edu.tw/cjlin/libsvm). This cross-classification scheme was applied separately within each background condition, within each attention condition and within each individual participant. The classification scheme was adapted to the specifics of the methods used here: it was applied per time point on the EEG data and per ROI in the fMRI data.

2.7.1. Time-resolved classification of location from EEG data

The time-resolved EEG classification analysis (Carlson et al., 2011b; Isik et al., 2014) served to determine the temporal dynamics with which category-independent location information emerged in the brain.

For each time point of the epoched EEG data, we extracted activations from 33 EEG channels. We chose the 33 central and posterior channels starting from the central midline, because we were interested in visual responses and previous studies had shown that location information was most pronounced in those areas (Graumann et al., 2022). We arranged activations from these channels into pattern vectors of 64 conditions and 60 raw trials. Raw trials were randomly arranged into four bins of 15 trials each and averaged by bin into four pseudo-trials to increase SNR. The classification procedure was repeated 100 times, each time assigning random trials into the bins before averaging into pseudo-trials. For classification, three of the pseudo-trials that came from two location conditions of the same category went into the training set. The model resulting from SVM classifier training was then tested on other pseudo-trials coming from the same two location conditions, but from a different category. The accuracy of the classification procedure was measured in percent classification accuracy (50% chance level). This amounted to 6 pairwise location classifications since we had 4 locations that were all classified pairwise once. During each iteration of pairwise location classification, the SVM was trained and tested across all combinations of the four categories in the training and testing set. For example, for a given location classification, the SVM was trained on faces and tested on animals (Fig. 3A). Then the same procedure was applied combing the remaining categories. With four categories in total, this resulted in 6 classification iterations to combine all categories into training and testing pairs. The direction of all training and testing pairs was reversed once (e.g., training on animals and testing on faces and vice versa), yielding a total of 12 classification iterations per pairwise location classification. We averaged 72 (6 location pairs × 12 category train/test pairs) classification accuracies in total per iteration. With 100 iterations with random trial assignment into pseudo-trials, this resulted in 7200 classification accuracies that were averaged per background condition, attention condition and participant. The result reflects the amount of location information that is independent of category at each time point, and within a background condition, attention condition and participant.

2.7.2. Time-resolved EEG searchlight in sensor space

To gain insights into which EEG channels contained the highest amount of location information we conducted a time-resolved EEG searchlight analysis in EEG channel space. This analysis followed the same scheme as the time-resolved EEG classification described above but extended it by one step: For each EEG channel c, the classification procedure was conducted not on all 33, but on the five closest channels surrounding c. The resulting classification accuracy was stored at the position of c. Iterating across all EEG channels with a temporal resolution downsampled to 10 ms steps, this yielded a map of classification accuracy across all channels and downsampled time points, for each participant, background condition and attention condition.

2.7.3. Time generalization analysis of location from EEG data

To characterize the neural dynamics of object location representations across time, we used temporal generalization analysis (Carlson et al., 2011b; Cichy et al., 2014; Isik et al., 2014; King and Dehaene, 2014).

In this analysis, the classification scheme was the same as in the time-resolved EEG classification but with the following extension: besides training and testing the SVM on data from the same time point, we additionally tested the SVM on data from all other time points within a −100 to 600 ms peristimulus time window, downsampled to a 10 ms temporal resolution. This resulted in a two-dimensional matrix of classification accuracies, indexed in rows and columns by the time points of data used for training and testing the SVM. This matrix indicates how much location information was shared at a given combination of time points. This analysis was conducted within time point combination, background condition, attention condition and participant.

2.7.4. Multivariate fMRI ROI analysis

The fMRI ROI classification analysis served to determine where category-independent location information emerged in the brain. For each ROI of the fMRI data, we extracted and arranged *t*-values into pattern vectors, one for each of the 16 conditions and 10 runs of the main experiment. Raw trials were randomly arranged into five bins with two runs each and averaged by bin into five pseudo-runs to increase SNR. We then proceeded with a 5-fold leave-one-pseudo-run-out-cross validation procedure. During each classification iteration, we trained an SVM on 4 and tested it on one pseudo-trial. The classification scheme was conceptually equivalent to the EEG classification. Training and testing was conducted across the two different categories, with each being in the training set once. We averaged across the two different training and testing directions of the two categories. The result reflects how much category-tolerant location information was present for each ROI, participant, background and attention condition separately.

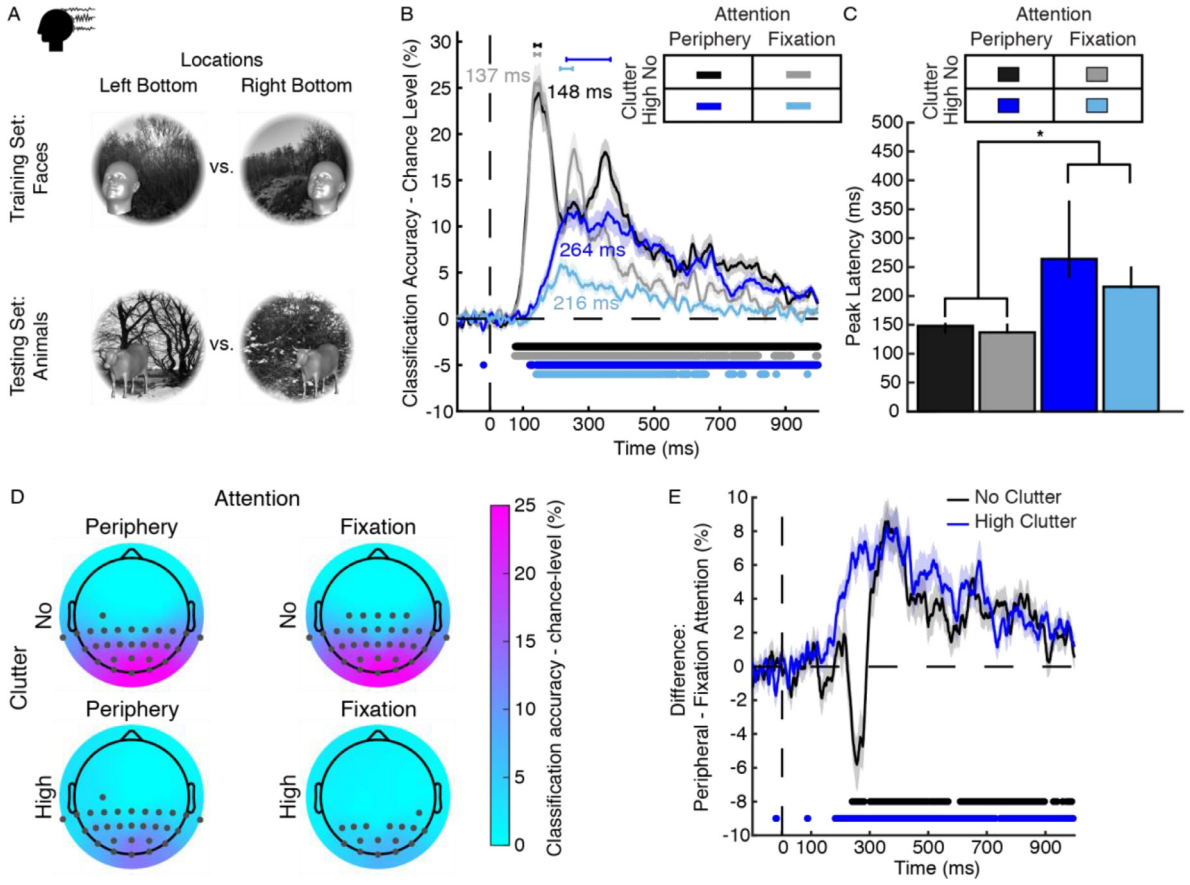

Fig. 3. Classification schemes and results of EEG location classification. **A,** Scheme for the classification of object location across categories within any background and attention condition. We trained a support vector machine (SVM) to distinguish brain activation patterns evoked by objects of a particular category presented at two locations (here: faces bottom left and right) and tested the SVM on activation patterns evoked by objects of another category (here: animals) presented at the same locations. Objects are enlarged here for display purposes. In the experiment objects did not extend across quadrants. **B,** Results of time-resolved location across category classification from EEG data. Results are color-coded by background and attention condition, with significant time points indicated by lines below curves ($N = 26$, $P<0.05$, FDR-corrected), 95% confidence intervals of peak latencies are indicated by lines above curves. Shaded areas around curves indicate SEM. **C,** Comparison of peak latencies of curves in B. Error bars represent 95% CIs. Stars indicate significant peak latency differences ($N = 26$, bootstrap test with 10,000 bootstraps). **D,** Results of the location across category classification searchlight in EEG channel space at peak latencies (as shown in B) in each condition. Significant electrodes are indicated by gray dots ($N = 26$, two-tailed Wilcoxon signed-rank test, $P<0.05$, FDR-corrected across electrodes and time points). **E,** Difference curves resulting from subtracting the time courses of the foveal from the peripheral attention condition in each background condition. Conventions as in B.

2.8. Statistical testing

2.8.1. Wilcoxon signed-rank tests

To test for above-chance classification accuracy at time points in the EEG time courses, in the EEG time-generalization matrix and for above-chance classification in the fMRI ROI results, we performed non-parametric two-tailed Wilcoxon signed-rank tests. The null hypothesis was always that the parameter being tested (i.e., classification accuracy) came from a distribution with a median of chance level (i.e., 50% classification accuracy for pairwise classification). We corrected the resulting *P*-values for multiple comparisons using false discovery rate at 5% level in every case where more than one test was conducted.

2.8.2. Bootstrap tests

To estimate confidence intervals and to compute the significance of peak-to-peak latency differences in the EEG time courses we used bootstrapping. We bootstrapped the participant pool 10,000 times with replacement and calculated the statistic of interest for each of the bootstrap samples.

For the peak-to-peak latency differences in the EEG time courses, we bootstrapped the latency difference between the peaks of the two time courses being compared. This resulted in a bootstrapped distribution that could be compared to zero. To determine the significance of peak-to-peak latencies in the EEG time courses, we computed the proportion of values that were equal to or smaller than zero and corrected them for multiple comparisons using FDR at $P = 0.05$. For computing the 95% confidence intervals of peak latencies of each time course, we bootstrapped the peak and computed the 95% percentiles of this distribution.

2.8.3. ANOVAs

We used repeated-measures ANOVAs to test for main effects and the interaction between the factors background and attention within ROIs. Since both factors had two levels, the assumption of sphericity was always met.

All post-hoc tests were conducted using pairwise *t*-tests and *P*-values were corrected for multiple comparisons using Tukey correction.

2.9. Code accessibility

The analysis code is publicly available via https://github.com/graumannm/AttentionLocation.

3. Results

For both the EEG and fMRI experiments the strategy to determine when and where attention modulates location representations was

equivalent: first we sought to establish H_R, i.e., that location representations of objects emerge at a later processing stage when objects are presented on cluttered backgrounds compared to blank backgrounds, independent of attention. On this basis we then arbitrated between H_D and H_S, i.e., whether attention dynamically modulates object location representations at different processing stages depending on background (H_D), or whether it statically modulated object location representations always at a late processing stage (H_S).

The difference between the EEG and the fMRI analyses lies in the way that the processing stages are determined: EEG determines the temporal delay with respect to image onset (Fig. 1A,B) and fMRI determines the region in the ventral visual stream (Fig. 1C,D) in which experimental effects emerge.

In the following we give the specifics of the EEG and fMRI experiments, the precise predictions, and the results. We begin with the EEG experiment determining the timing of attentional modulation, followed by the fMRI experiment determining where in the visual processing hierarchy the attentional modulation occurs.

3.1. Attentional modulation of object location representations in time

For the EEG experiment we used an experimental design with fully crossed experimental factors background and attention with two levels per factor (2 background condition levels × 2 attention condition levels, Fig. 2A).

In detail, participants saw objects from four different categories, each presented in four locations (Fig. 2A). The two background conditions were no and high clutter. Each object in each location was presented in both background conditions. Each combination of object category, location and background was then also crossed with the two levels of attention conditions: the peripheral and fixation attention conditions (Fig. 2A). Attention conditions were solely defined by the task that participants performed, while visual stimulation was identical (Fig. 2E). In the peripheral attention condition, participants directed their covert spatial attention to the periphery and responded to a catch object (glass) with button press (Fig. 2A,E). In the fixation attention condition, participants performed a demanding task on fixation to remove their spatial attention from the objects in the periphery (Fig. 2A,E). Overall, participant's performance was high in all conditions (Supplementary Fig. 1), with higher performance in the fixation than in the peripheral attention task (for details see Supplementary Fig. 1).

In total, the 2 × 2 experimental design resulted in 4 factor combinations. We performed a time-resolved and pair-wise classification analysis of location across category within each of these four factor combinations separately (Fig. 3A,B). This meant training a classifier to distinguish between millisecond-specific EEG pattern vectors associated with two locations and testing on a held-out testing data set associated with the same two locations. We performed the classification across object category, that is training on data associated with locations from one object category and testing on data from another category (Fig. 3A). This ensured that location classification results were not confounded with category information and allowed us to draw conclusions about location representations independent of object category representations.

3.1.1. The temporal dynamics of object location representations with blank and cluttered backgrounds

To lay the basis for later analyses on attentional modulation, we first tested H_R, i.e., that location representations of objects with clutter emerge later than on blank backgrounds, independent of attention. For this we determined and compared the latencies of the classification peaks in the EEG time courses of both background conditions, assuming that the peaks represent the time points at which representations become most differentiable (DiCarlo and Cox, 2007). Our prediction was that location information would peak later in the high than in the no clutter condition, because dissecting objects from the background requires additional grouping and segmentation operations implemented in recurrent processing and thus increasing processing time (Groen et al., 2018; Seijdel et al., 2020, 2021; Graumann et al., 2022).

The results of the time-resolved location classification are shown in Fig. 3B. We read out location information from the EEG signal in all background and attention conditions above chance level ($N = 26$, two-tailed Wilcoxon signed-rank test, $P<0.05$, FDR-corrected across time points).

Focusing on peak latencies (95% confidence intervals reported in brackets, $N = 26$, 10,000 bootstrap samples), we observed that time courses in the no and high clutter conditions peaked at different times. In the no clutter condition, location information peaked early, regardless of attention condition (Fig. 3B; peak latency peripheral condition: 148 ms (135–153.5 ms); peak latency fixation condition: 137 ms (135–152 ms)). With high clutter, location information peaked later in both attention conditions (Fig. 3B; peripheral condition: 264 ms (232–365 ms); fixation condition: 216 ms (213–251 ms)). Onset latencies followed a similar pattern, with significant classification onsets being earlier in the no clutter conditions (peripheral condition: 77 ms; fixation condition: 75 ms) than in the high clutter conditions (peripheral condition: 121 ms; fixation condition: 140 ms). To test whether the peak latencies across background conditions were significantly different, we bootstrapped the peak-to-peak latency differences between pairs of no and high clutter condition peaks (Fig. 3C, 95% confidence intervals in brackets, $N = 26$, bootstrap test, 10,000 bootstraps, FDR-corrected). This was done both within and across attention conditions. Overall, the results clearly and consistently support H_R (delayed emergence of location representations with clutter compared to no clutter, independent of attention). Location information peaked significantly earlier in the no compared to the high clutter conditions independent of attention condition: Within attention condition, the peak-to-peak latency difference between background conditions was 116 ms (83–223 ms; $P<0.001$) in the peripheral attention condition and 79 ms in the fixation attention condition (63–114 ms; $P<0.001$). Across attention conditions, the delays between background condition peaks were also significant (peripheral attention and no clutter condition vs. fixation attention and high clutter condition: 68 ms delay, 63–105 m; $P<0.001$; fixation attention and no clutter condition vs. peripheral attention and high clutter condition: 127 ms delay, 83–224 ms, $P<0.001$). Following the results of three control analyses (Supplementary Fig. 2), we assessed the contribution of confounding eye-movement artefacts to be unlikely.

Additional analyses of the observed effects reproduced previously observed characteristics of object location representations (Graumann et al., 2022) and thus further supported H_R. A searchlight analysis in EEG sensor space (Fig. 3D) localized the sources of the peaks to occipito-temporal electrodes (Fig. 3D), suggesting the locus of object location representations to be in occipital and temporal cortices. A supplementary time-generalization analysis (King and Dehaene, 2014) showed that location representations for objects on blank and cluttered backgrounds emerged within the same processing stage, but with a delay with cluttered backgrounds (Supplementary Fig. 3, Supplementary Methods 1).

Interestingly, we also observed an unexpected result: in the no clutter condition, the latency of the 2nd peak was earlier in the fixation attention condition (256 ms) than in the peripheral attention condition (350 ms). Supplementary analyses revealed that location information in the two no clutter conditions was shared between the two 2nd peaks (Supplementary Fig. 4 A,B) with comparable topographies (Supplementary Fig. 4 C,D). Such shared information across time has in the past been linked to delays due to recurrence (Graumann et al., 2022) and might in this case be related to top-down attentional modulation implemented in long-range feedback from prefrontal areas (Squire et al., 2013) requiring additional processing time.

Together, these results provide empirical evidence for the hypothesis H_R, that location representations of objects with clutter emerge later than on blank backgrounds, independent of attention.

3.1.2. Late attentional modulation of location representations independent of background

Affirming H_R formed the basis for arbitrating between our main hypotheses H_D and H_S. H_D predicts that attentional modulation is highest when location information is highest: with no clutter, it predicts an early modulation in time of location representations and with high clutter it predicts a late modulation in time (Fig. 1B). H_S states that spatial attention modulates location representations always late, after the end of the bottom-up response at ~100–150 ms (Lamme and Roelfsema, 2000; VanRullen and Thorpe, 2001; Fahrenfort et al., 2007; Camprodon et al., 2010; Koivisto et al., 2011). Thus, H_D predicts an interaction between attention and background and H_S predicts that they are independent.

To assess H_S and H_D we determined the time course of attentional modulation in both background conditions. Attentional modulation was defined as an enhancement of representations (Desimone and Duncan, 1995; Reynolds and Chelazzi, 2004; Briggs et al., 2013). To quantify attentional modulation, we subtracted classification accuracies in the fixation attention condition from the peripheral attention condition, within each background condition. Since visual stimulation was identical across attention conditions, we attributed differences between them to attentional modulation.

Fig. 3E shows the result of this analysis. We found attentional modulation of location representations in both background conditions in a late time window, providing clear evidence for H_S. In detail, we observed a significant positive difference in the no clutter condition starting from 301 ms, reflecting attentional modulation (Fig. 3E; $N = 26$, two-tailed Wilcoxon signed-rank test, $P<0.05$, FDR-corrected across time points). In the high clutter condition, we found evidence for attentional modulation starting from 182 ms (Fig. 3E; $N = 26$, two-tailed Wilcoxon signed-rank test, $P<0.05$, FDR-corrected across time points), as reflected in a significant positive difference that lasted until the end of the time window.

Although attentional modulation occurred in the late time window (>150 ms) in both background conditions, the significance onset of attentional modulation was earlier in the high (182 ms) than in the no clutter condition (240 ms), indicating a slightly earlier effect of attention on location representations of objects on cluttered backgrounds than on blank backgrounds. This indicates that neural responses to cluttered images are sensitive to attentional effects earlier than with blank backgrounds.

Together, these results constitute strong evidence for H_S, showing that attention modulates object location representations in a late time window after the bottom-up response, independent of background.

3.1.3. Dissecting transient and persistent components of attentional modulation

While clearly supporting H_S (late attentional modulation independent of background), the results hitherto do not yet characterize the temporal dynamics underlying attentional modulation of location representations. Typically during visual perception, time-resolved multivariate results reflect a conglomerate of both rapidly changing transient information flow as well as persistent activity which maintains certain types of information over long stretches of time (Cichy et al., 2014; King and Dehaene, 2014).

Thus, here we investigated whether attention and background modulate persistent, transient or both aspects of location representations. For this we conducted temporal generalization analysis (King and Dehaene, 2014). This resulted in two-dimensional time generalization matrices, indexed in both dimensions in time indicating similarities of object location representation across time. While transient representations are reflected as high information on the diagonal of such matrices, persistent representations are found off-diagonal.

As previously, we classified location representations within background and attention condition, resulting in 4 time-generalization matrices (Fig. 4A,B,D,E), corresponding to the 4 classification time courses above (Fig. 3B). We first present the single results ordered by background condition, before quantifying attentional modulation.

In the no clutter condition, we found similar results in both attention conditions (Fig. 4A,B): location information peaked early at ~100 ms on the diagonal, representing transient information flow ($N = 27$, $P<0.05$, two-tailed Wilcoxon signed-rank test, FDR-corrected). Starting from ~250 ms, location information generalized more broadly across time points, indicating persistent information. In the high clutter condition (Fig. 4D,E) information generalized broadly across time points starting from ~140 ms in both attention conditions, ($N = 27$, $P<0.05$, two-tailed Wilcoxon signed-rank test, FDR-corrected), indicating persistent information. Transient information peaked on the diagonal starting from ~240 ms.

We quantified attentional modulation as above (Fig. 3E) by comparing the classification results for the two attention conditions, subtracting the results of the fixation attention condition from the peripheral attention condition.

We found that spatial attention modulated both transient and persistent representations in late time windows, independent of background. In the no clutter condition, attention modulated the persistent clusters from ~230 ms and both transient and persistent information from ~300 ms (Fig. 4C; $N = 27$, $P<0.05$, two-tailed Wilcoxon signed-rank test, FDR-corrected). In the high clutter condition, spatial attention modulated location representations across the entire time window starting from ~180 ms (Fig. 4F; $N = 27$, $P<0.05$, two-tailed Wilcoxon signed-rank test, FDR-corrected). Although overall attentional modulation was observed in a late time window (>150 ms) in both background conditions, we observed an earlier effect of attentional modulation in the high clutter condition than in the no clutter condition, similar to the time course results (Fig. 3E). This might indicate that the noisy input created by cluttered backgrounds renders the neural responses to these stimuli more sensitive to attentional modulation.

In sum, we found that attention modulates both transient and persistent representations of object location in late time windows beyond 150 ms.

3.2. Clutter and attention independently affect location representations along the ventral visual stream

We proceed to investigate which visual processing stages are modulated by background and attention in an fMRI experiment, determining processing stages by localizing and assessing cortical regions of the ventral visual stream (Fig. 1C,D). In this context H_R predicts that location representations of objects emerge in higher regions along the ventral stream when objects are presented on cluttered backgrounds compared to blank backgrounds (Graumann et al., 2022; Fig. 1C). H_D predicts that attentional modulation is high where location information is high (Fig. 1D): with no clutter, attention modulates location representations throughout the ventral stream and with high clutter attention modulates location representations in mid- or high-level visual areas. H_S instead predicts that attentional modulation is high in mid- and high-level visual areas only, independent of background.

We further investigated the dorsal cortex because it is assumed to process visuospatial information (Ungerleider and Haxby, 1994; Milner and Goodale, 2006; Kravitz et al., 2011; Groen et al., 2022) and it has also been implicated in attentional processing (Silver et al., 2005; Szczepanski et al., 2010; Sprague and Serences, 2013). In contrast to those findings, more recent work did not find evidence for strong representations of object location in the dorsal stream (Graumann et al., 2022). However, that study did not manipulate attention. To investigate whether location representations in the dorsal stream depend on attention and resolve inconsistencies in previous research, we also included regions from the dorsal stream in our analyses.

The design of the fMRI experiment was equivalent to the design in the EEG experiment with a reduced number of levels for the factors cate-

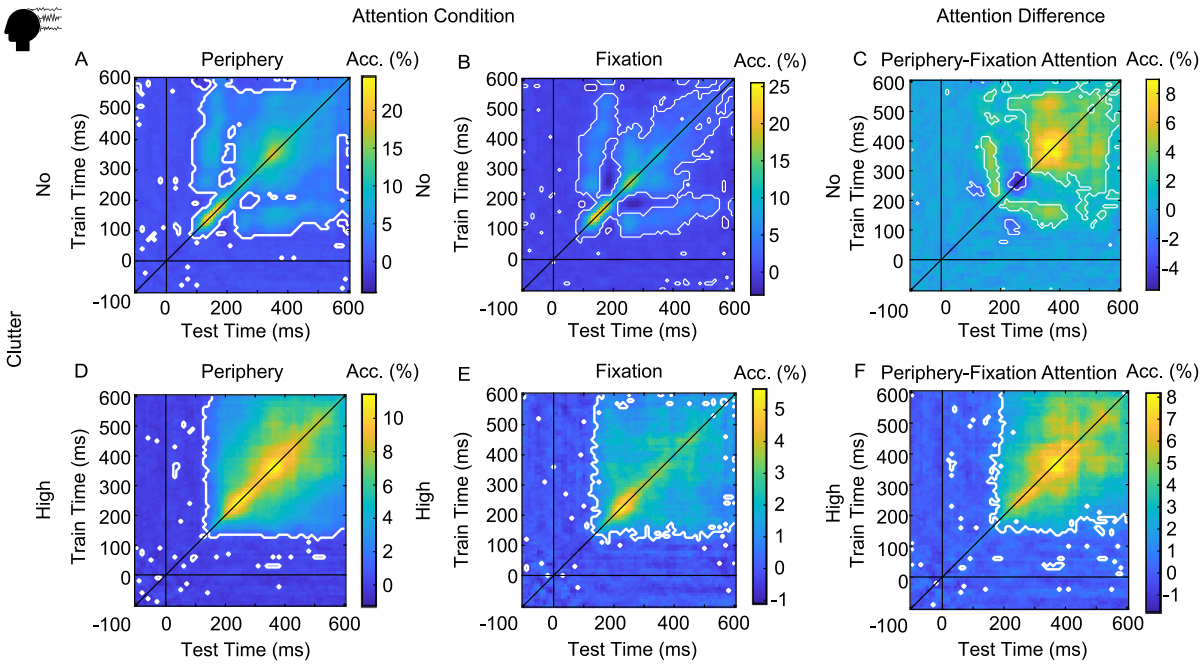

Fig. 4. EEG results of time-generalization analyses within each background and attention condition. Rows represent background and columns represent attention conditions. **A,** Location classification across categories and time points in the no clutter & peripheral attention condition. Horizontal and vertical black lines indicate stimulus onset, oblique black line highlights the diagonal. White outlines indicate significant time points (N = 26, two-tailed Wilcoxon signed-rank test, P<0.05, FDR-corrected). **B,** Location classification across categories and time points in the no clutter & fixation attention condition. **C,** Difference matrix resulting from subtracting the matrices representing fixation (B) from peripheral attention (A) in the no clutter condition. Plot conventions as in A. **D,** Location classification across categories and time points in the high clutter & peripheral attention condition. **E,** Location classification across categories and time points in the high clutter & fixation attention condition. **F,** Difference matrix resulting from subtracting the matrices representing fixation (E) from peripheral attention (D) in the high clutter condition.

gory and location. This adaptation was made to accommodate the longer trial duration required for our fMRI event-related design (Fig. 2D) while maintaining a feasible session duration. We presented objects from two categories (faces, cars) in two locations (left and right horizontally from fixation; Fig. 2B) instead of four categories and locations. To characterize the effects of background and spatial attention on location representations in visual cortex, we defined regions-of-interest (ROIs) along the ventral stream, since H_R predicted effects to emerge there based on previous studies (Hong et al., 2016; Graumann et al., 2022). We also included ROIs along the dorsal stream to resolve contradictory findings from previous research.

We classified location across category using an analogous classification scheme as in the EEG experiment. We trained a classifier on fMRI patterns associated with two locations of one object category and subsequently cross-validated the classifier on new testing data associated with the same locations of a new object category. This classification was performed in each ROI separately, for each level of the background condition (no clutter, high clutter) and for each level of attention conditions (periphery, fixation) separately, resulting in four classification accuracies per ROI and subject. We included three ROIs in early visual cortex (V1, V2, V3), two ROIs in the ventral stream (V4, LOC) and four ROIs in the dorsal stream (IPS0, IPS1, IPS2, SPL).

We tested H_R, H_S and H_D in 2 × 2 repeated-measures ANOVAs (N = 20, FDR-correction for multiple comparisons) with factors background (no clutter, high clutter) and attention (peripheral, fixation) in all ROIs of the ventral and dorsal visual streams, focusing on the ventral visual stream first.

Hypothesis H_R predicted a main effect of background in early visual areas, but not in high-level visual areas of the ventral visual stream. Consistent with the predictions of H_R, we found significant main effects of background in early visual areas V1 and V2 (Fig. 5A,B; V1: $F_{(1,19)}$=9.88, P = 0.005, partial η^2 = 0.34; V2: $F_{(1,19)}$=11.56, P = 0.003, partial η^2 = 0.3), but not in mid- and high-level visual areas V3 and LOC (Table 1), except for V4, which also showed a main effect of background ($F_{(1,19)}$=16.64, P<0.001, partial η^2 = 0.47). In line with this, a supplementary ANOVA comparing the background difference between ROIs additionally revealed a significant main effect of ROI with higher background differences in V2 and V4 than in LOC (Supplementary Fig. 5A).

On this basis arbitrating between H_S and H_D we found clear evidence for H_S (the hypothesis predicting late attentional modulation independent of background). Location information in mid- and high-level ventral visual areas V3, V4 and LOC all showed a significant main effect of attention (Fig. 5C,D,E; V3: $F_{(1,19)}$=13.36, P = 0.002, partial η^2 = 0.41; V4: $F_{(1,19)}$=45, P<0.001, partial η^2 = 0.70; LOC: $F_{(1,19)}$=24.04, P<0.001, partial η^2 = 0.56), but no significant interaction between background and attention as would have been predicted by H_D. These results were confirmed through a univariate analysis that revealed main effects of clutter in early- and mid-level areas V1, V2, V3 and V4 and main effects of attention in mid- and high-level areas V4 and LOC (Supplementary Fig. 6A). Comparing attentional modulation across ROIs provided further evidence for H_S rather than H_D: attention differences across ROIs showed a significant main effect of ROI with higher attentional modulation in LOC than in V2 (Supplementary Fig. 5B). However, univariate activation differences between attention conditions along the ventral stream were significantly higher in the high than in the no clutter condition (Supplementary Fig. 6B), but we found no evidence for a significant univariate effect of attention in V1. Furthermore, univariate results showed a right-lateralized effect of clutter in V1 and V4 and a left lateralized effect in V2 (Supplementary Fig. 6C,D). These analyses were exploratory and future studies are needed to see if they replicate.

Equivalent testing in the dorsal visual stream revealed no significant main or interaction effect in any regions along the dorsal stream

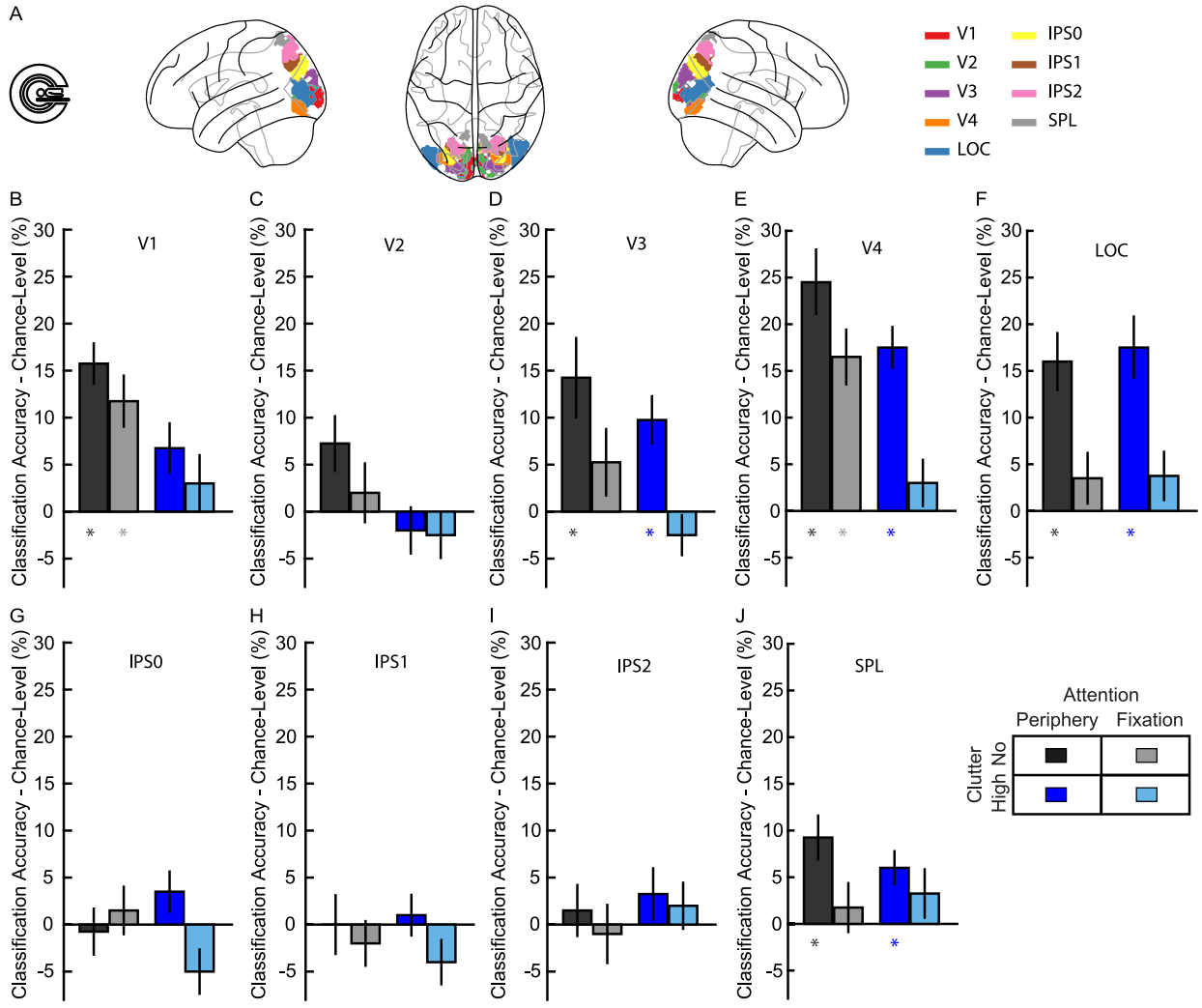

Fig. 5. Location across category classification results in the four conditions in early (V1, V2, V3), ventral (V4, LOC) and dorsal (IPS0–2, SPL) visual ROIs. Stars below bars indicate significant above-chance classification (N = 20, two-tailed Wilcoxon signed-rank test, P<0.05 false discovery rate (FDR) corrected). Error bars represent standard error of the mean (SEM). A, ROIs on cortical surface. B, V1. C, V2. D, V3. E, V4. F, LOC. G, IPS0. H, IPS1. I, IPS2. J, SPL.

Table 1
Results of the 2 × 2 repeated-measures ANOVA (N = 20) with factors background (no clutter, high clutter) and attention (peripheral, fixation), analyzing location classification accuracies in 9 ROIs that were included in the analyses. Asterisks behind P-values indicate significance with FDR correction across number of comparisons (for 9 ROIs).

ROI	Main effect background	Main effect attention	Interaction effect
V1	$F_{(1,19)}$=9.88, P = 0.005*, partial η^2 = 0.34	$F_{(1,19)}$=2.30, P = 0.15, partial η^2 = 0.11	$F_{(1,19)}$=0.00, P = 0.97, partial η^2 = 0.00
V2	$F_{(1,19)}$=11.60, P = 0.003*, partial η^2 = 0.38	$F_{(1,19)}$=0.94, P = 0.35, partial η^2 = 0.38	$F_{(1,19)}$=0.58, P = 0.46, partial η^2 = 0.03
V3	$F_{(1,19)}$=3.48, P = 0.08, partial η^2 = 0.16	$F_{(1,19)}$=13.36, P = 0.002*, partial η^2 = 0.41	$F_{(1,19)}$=0.67, P = 0.42, partial η^2 = 0.03
V4	$F_{(1,19)}$=16.64, P<0.001*, partial η^2 = 0.47	$F_{(1,19)}$=45, P<0.001*, partial η^2 = 0.70	$F_{(1,19)}$=2.21, P = 0.15, partial η^2 = 0.10
LOC	$F_{(1,19)}$=0.08, P = 0.78, partial η^2 = 0.00	$F_{(1,19)}$=24.04, P<0.001*, partial η^2 = 0.56	$F_{(1,19)}$=0.06, P = 0.81, partial η^2 = 0.00
IPS0	$F_{(1,19)}$=0.29, P = 0.60, partial η^2 = 0.02	$F_{(1,19)}$=1.40, P = 0.25, partial η^2 = 0.07	$F_{(1,19)}$=5.45, P = 0.03, partial η^2 = 0.22
IPS1	$F_{(1,19)}$=0.03, P = 0.87, partial η^2 = 0.00	$F_{(1,19)}$=2.41, P = 0.14, partial η^2 = 0.11	$F_{(1,19)}$=0.53, P = 0.47, partial η^2 = 0.03
IPS2	$F_{(1,19)}$=0.97, P = 0.34, partial η^2 = 0.05	$F_{(1,19)}$=0.58, P = 0.46, partial η^2 = 0.03	$F_{(1,19)}$=0.03, P = 0.87, partial η^2 = 0.00
SPL	$F_{(1,19)}$=0.12, P = 0.74, partial η^2 = 0.01	$F_{(1,19)}$=4.38, P = 0.05, partial η^2 = 0.19	$F_{(1,19)}$=0.72, P = 0.41, partial η^2 = 0.04

(Fig. 5F,G,H,I; Table 1), consistent with the observation that object location representations emerge rather along the ventral than the dorsal stream (Hong et al., 2016; Graumann et al., 2022).

In sum, the results of the fMRI experiment concur with the results of the EEG experiment in providing evidence for H_R and H_S: object location representations emerge gradually along the processing hierarchy of the ventral visual stream in line with H_R, and attention modulates object location representations in mid- and high-level ventral areas independent of the object's background, in line with H_S.

4. Discussion

Using EEG and fMRI we investigated at which stage of the visual processing hierarchy attention modulates object location representations. Our results converge across the two experiments and imaging modalities into a common view. We reproduced the recent observation that object location representations emerge at later processing stages when presented on cluttered than on blank backgrounds (H_R) and showed that this holds independent of attention. On this basis we examined the

effect of attention on object location representations, finding that attention modulated location representations statically during late stages of visual processing in cortical time and space, independent of the object's background (H_S).

4.1. Disentangling the influences of background and attention on the temporal dynamics of location representations

Recent research has revealed that object location representations emerge later in the ventral visual processing hierarchy when objects appear on cluttered rather than on blank backgrounds (Hong et al., 2016; Graumann et al., 2022). However, it remained unclear to which degree this effect relied on or was influenced by attention or not. Previous research has highlighted attention as important for object perception under cluttered conditions (Treisman and Gelade, 1980; Wolfe, 1994; Reddy and Kanwisher, 2007; Lee and Maunsell, 2010). Further, both temporal delays observed for object perception and attention have been related to recurrent processes (Tang et al., 2014; Kar et al., 2019; Rajaei et al., 2019; van Bergen and Kriegeskorte, 2020), suggesting shared neural mechanisms.

Here, we clarify the relationship and find a dissociation: while cluttered viewing conditions delay processing (see also time-generalization analysis in Supplementary Fig. 3, Supplementary Methods 1), spatial attention in contrast increases information without changing its timing. These results suggest that background clutter and attention have differential effects on object location processing. Background clutter, like other factors that increase image complexity, triggers local recurrent processes that can be measured in delayed responses (Tang et al., 2014, 2018; Groen et al., 2018; Kar et al., 2019; Rajaei et al., 2019; Seijdel et al., 2021; Graumann et al., 2022). These effects might be driven by the average contrast of a cluttered scene, which increases the need for segmentation operations (Groen et al., 2018; Seijdel et al., 2020, 2021). In contrast, spatial attention triggers modulation of neural responses that can be measured as enhancement of response magnitude (Desimone and Duncan, 1995; Reynolds and Chelazzi, 2004; Briggs et al., 2013).

However, an unexpected result in the no clutter condition indicates a potential interaction between recurrence and attention. In both no clutter conditions, after the initial peak at ~140 ms, we observed second peaks. These second peaks might represent a recurrent processing loop, while the first peaks mark the feedforward sweep (Kietzmann et al., 2019). Unexpectedly, the second peak with attention on periphery was delayed compared to the second peak with attention on fixation. This delay might represent long-range feedback from prefrontal areas implementing attentional modulation, requiring additional processing time. However, our supplementary analyses cannot elucidate whether this is the underlying mechanism and thus this result needs further investigation in future research.

4.2. Attention modulates location representations later than the initial bottom-up response

The EEG results revealed attentional modulation of location representations in a late time window beyond the first 150 ms of the bottom-up response, independent of the object's background. This directly supports the hypothesis H_S i.e., that the processing stage of attentional modulation is statically late and refutes the hypothesis H_D, i.e., that the processing stage of attentional modulation changes dynamically depending on the background. In-depth investigation further revealed attentional modulation of transient and persistent temporal dynamics of location representations. This modulation was likewise found in late time windows independent of background, providing further evidence for H_S.

Interestingly, we also observed that the onset of attentional modulation for location representations of objects with high clutter, although late (>150 ms), occurred earlier than the onset of attentional modulation when objects were presented on blank backgrounds. This might be related to increased noise added to the visual information by the background clutter which might render neural responses sensitive to attentional fine-tuning earlier in time than with blank backgrounds.

Our results are seemingly at odds with earlier studies finding attentional modulation before 150 ms in the P1 (Hillyard et al., 1998b; Luck et al., 2000; Itthipuripat et al., 2019) and the N1 (Mangun, 1995; Hillyard et al., 1998a; Itthipuripat et al., 2019) component. How is this discrepancy to be explained? One possible reason might be that above mentioned studies used space-based attention where participants were cued to attend either to the left or right whereas our study uses a difference in spatial deployment of attention that interacts with the incoming stimulus. It is possible that pre-stimulus cueing on the left or right side of fixation, as used in earlier paradigms, bundles attentional resources earlier and stronger, resulting in early attentional effects on ERPs. In comparison, our study directed attention towards the periphery which might have diluted attention more compared to previous work. Another possible reason might be the choice of the stimulus in general. Above mentioned ERP studies employed simple artificial stimulation conditions which might elicit attentional modulation already early. However, later studies using naturalistic stimuli, comparable to the ones used here, did not find early attentional modulation (VanRullen and Thorpe, 2001; Groen et al., 2016; Kaiser et al., 2016a; Battistoni et al., 2020). Together this questions the degree to which previously observed effects of early attentional modulation generalize to more complex stimuli and naturalistic viewing conditions encountered in the real world where the location of attended objects is not always predictable.

Another contributing factor to the discrepancy could be that attentional enhancement of early neural responses is stronger when the visual task is more difficult or when visual processing is overloaded (Spitzer et al., 1988; Lavie, 1995; Luck et al., 2000; Boudreau et al., 2006; Chen et al., 2008) which might have been the case in earlier ERP studies e.g. by presenting stimuli in faster sequences (Hillyard et al., 1998b). In contrast in our experiment, stimuli in the no clutter condition were highly salient and presented long enough to be clearly visible. Future research comparing attentional modulation of artificial vs. real-world stimuli with different levels of task difficulty are needed to resolve this issue. In our study, differences in task difficulty might in turn have strengthened the attentional manipulation: our behavioural results suggest that the peripheral task was harder than the fixation task in the EEG experiment (Supplementary Fig. 1), possibly due to multiple vs. single target locations. The attentional effects might thus haven been enhanced by task difficulty (Boudreau et al., 2006; Chen et al., 2008).

We cannot rule out that participants overtly or covertly explored the image in the high clutter condition. However, we believe that such effects should be minor or absent, because objects were presented close to fixation (4,2° of visual angle from fixation to object center), rendering them clearly visible from fixation at all times. Furthermore, such effects should introducte noise and therefore attenuate the attentional modulation in the high clutter compared to no clutter condition with attention on periphery, which is not what we observe in our results (Fig. 3E). Finally,object location was classified as early as 120 and 140 ms with high clutter, which is well before the onset of saccades at ~200–250 ms (Carrasco et al., 2011), making overt exploration unlikely.

4.3. Attentional modulation in mid- and high-level ventral visual areas

Consistent with the EEG results indicating attentional modulation of later visual object processing stages in time, the fMRI experiment localized those modulations to mid- and high-level ventral visual areas. While our results do not exclude the existence of attentional modulation also in early visual cortex as observed previously (Roelfsema et al., 1998; Gandhi et al., 1999; Martínez et al., 2001; Noesselt et al., 2002; Khayat et al., 2006; Lakatos et al., 2008; Briggs et al., 2013; Herrero et al., 2013; Itthipuripat et al., 2019), they suggest that the

modulation might be strongest and thus most likely to be detected at later stages of ventral visual cortex (Murray and Wojciulik, 2004; Buffalo et al., 2010; Peelen and Kastner, 2014; Kay et al., 2015).

Like previous studies, we found that high-level ventral visual cortex encodes retinotopic location representations (Cichy et al., 2011; Golomb and Kanwisher, 2012). It might seem surprising that we do not find more prominent location representations in the dorsal stream, given that it has traditionally been labelled the "where" pathway (Ungerleider and Haxby, 1994) and has been related to visuospatial processing. However more recently, dorsal stream representations have been closely linked to vision for goal-directed behavior and are modulated by task (Bracci et al., 2017; Vaziri-Pashkam and Xu, 2017; Hebart et al., 2018). Therefore, the dorsal stream might represent action-relevant visual information rather than visuospatial information per se (Milner and Goodale, 2006).

Our results add further evidence towards the view that attentional modulation begins in higher processing stages and is then relayed back to lower stages (Buffalo et al., 2010), which could be reflected as increasing attentional modulation along the ventral stream (Kay et al., 2015).

4.4. Location representations of objects on cluttered backgrounds in the ventral stream

The fMRI results reveal a double dissociation between the effects of clutter and attention on early and late ventral visual areas: early visual areas show an effect of background but not of attention, while the reverse is true for mid- and high-level visual areas. Put differently, we find that both robustness to clutter (Hong et al., 2016; Graumann et al., 2022) and attentional modulation increase along the ventral visual stream (Buffalo et al., 2010; Kay et al., 2015). We speculate that these phenomena depend on the common mechanistic and computational basis of receptive field size increases along the ventral visual stream. Attention increases population receptive field (pRF) size in higher-level ventral areas, thereby enhancing location sensitivity (Kay et al., 2015). In addition, the biased competition model of attention predicts that if an object appears on a cluttered background, attention biases the neural response towards the relevant object, while suppressing the response to the background clutter within the same RF (Kastner and Ungerleider, 2001; Reddy and Kanwisher, 2007; Reddy et al., 2009). Together, these computational models and previous findings might explain why both attentional modulation and location information increase along the ventral visual stream.

An increase in pRF size might simultaneously benefit object segmentation from cluttered backgrounds by encoding object location in global voxel patterns (Eurich and Schwegler, 1997; Kay et al., 2015). This benefit for object segmentation and biased competition within RFs might in contrast not be present in early visual cortex where RF size is small (Wandell and Winawer, 2015) and cells respond in a location-unspecific way across all stimulated portions of the visual field to both objects and background clutter.

4.5. Limitations

We highlight three limitations of our experimental designs that are important for the correct interpretation of the results.

The first limitation is that in our experiment object locations and the content of the background are randomly paired and thus incongruent. In contrast, in the real world objects typically appear in locations congruent with the background scene. Attentional selection can exploit such relations between objects and backgrounds (Wolfe et al., 2011; Kaiser et al., 2019; Võ et al., 2019; Battistoni et al., 2020) on the basis of scene gist information (Oliva, 2005; Greene and Oliva, 2009). In our experiment this type of information cannot be exploited. Thus, when object locations and scene background are congruent, attentional modulation might be faster than revealed here. The flipside of the limitation is that our experimental design isolates the effect of clutter on visual processing and attentional modulation independent of congruency effects. To determine the effect of congruency of object location and background on visual processing, studies are needed that additionally investigate congruency as an experimental factor.

The second limitation is that we did not directly assess the behavioral effects of attentional modulation on localization performance. Spatial attention benefits object localization in cluttered displays (Treisman and Gelade, 1980; Wolfe, 1994; Wolfe et al., 2011) by increasing processing speed. Future studies may combine assessment with brain imaging to link the effect of attention for objects on cluttered backgrounds in brain and behavior.

Our results might be explained not solely by the effect of spatial attention, but might be influenced by task difficulty, task relevance and feature-based attention, too. Behavioural data suggest that the peripheral task was more difficult than the fixation task (Supplementary Fig. 1) because hit rates were higher in the fixation than in the peripheral task. Task difficulty enhances effects of spatial attention (Boudreau et al., 2006) by enhancing responses at the attended locations and suppressing responses at non-attended locations. Therefore, the effect of spatial attention in the peripheral task might have been boosted by task difficulty. Further, task relevance varied along with attention in our experiment. However we do not expect this to drive our results, because task relevance shows a stronger effect in dorsal than in ventral regions (Hebart et al., 2018; Vaziri-Pashkam and Xu, 2017; Bracci et al., 2017), which is contrary to our results. Finally, our task manipulated not only spatial attention (fixation vs. periphery) but also feature-based attention, because digits and objects vary in spatial frequency (high for digits, low for objects). But since all of our analyses were carried out on objects in the periphery, spatial attention was most likely the main driver of our results.

4.6. Conclusion

In daily life, we use our spatial attention to help us focus on relevant portions of the visual field in cluttered environments (Wolfe et al., 2011). Our results clarify that attention modulates object location representations at late processing stages, using both spatial and temporal markers. Furthermore, they establish that attentional modulation is a cognitive process which is separate from recurrent processes which are engaged when objects appear in cluttered environments.

Declaration of Competing Interest

None.

Credit authorship contribution statement

Monika Graumann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Lara A. Wallenwein:** Data curation. **Radoslaw M. Cichy:** Conceptualization, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing.

Data availability

The fMRI and EEG data are publicly available via https://osf.io/hf6zp/.

Acknowledgements

We thank Benjamin Lahner for the glass brain plots. Computing resources were provided by the high-performance computing facilities at ZEDAT, Freie Universität Berlin. EEG and fMRI data were acquired at the Center for Cognitive Neuroscience, Freie Universität Berlin, Berlin. M.G. and R.M.C. are

supported by German Research Council (DFG) (CI241/1-1, CI241/3-1, CI241/7–1). R.M.C. is supported by the European Research Council (ERC-StG-2018-803370). L.A.W. is supported by the University of Konstanz. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2023.120053.

References

Battistoni, E., Kaiser, D., Hickey, C., Peelen, M.V., 2020. The time course of spatial attention during naturalistic visual search. Cortex 122, 225–234.
Boudreau, C.E., Williford, T.H., Maunsell, J.H.R., 2006. Effects of task difficulty and target likelihood in area V4 of macaque monkeys. J. Neurophysiol. 96, 2377–2387.
Bracci, S., Daniels, N., De Beeck, H.O., 2017. Task context overrules object- and category-related representational content in the human parietal cortex. Cereb. Cortex 27, 310–321.
Briggs, F., Mangun, G.R., Usrey, W.M., 2013. Attention enhances synaptic efficacy and the signal-to-noise ratio in neural circuits. Nature 499, 476–480.
Buffalo, E.A., Fries, P., Landman, R., Liang, H., Desimone, R., 2010. A backward progression of attentional effects in the ventral stream. Proc. Natl Acad. Sci. 107, 361–365.
Camprodon, J.A., Zohary, E., Brodbeck, V., Pascual-Leone, A., 2010. Two phases of V1 activity for visual recognition of natural images. J. Cogn. Neurosci. 22, 1262–1269.
Carlson, T.A., Hogendoorn, H., Fonteijn, H., Verstraten, F.A., 2011a. Spatial coding and invariance in object-selective cortex. Cortex 47, 14–22.
Carlson, T.A., Hogendoorn, H., Kanai, R., Mesik, J., Turret, J., 2011b. High temporal resolution decoding of object position and category. J. Vis. 11, 1–17.
Carrasco, M., 2011. Visual attention: the past 25 years. Vision Res. 51, 1484–1525..
Chang, C.-.C., Lin, C.-.J., 2011. Libsvm: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 1–27.
Chaumon, M., Bishop, D.V.M., Busch, N.A., 2015. A practical guide to the selection of independent components of the electroencephalogram for artifact correction. J. Neurosci. Methods 250, 47–63.
Chen, Y., Martinez-Conde, S., Macknik, S.L., Bereshpolova, Y., Swadlow, H.A., Alonso, J.M., 2008. Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. Nat. Neurosci. 11, 974–982.
Cichy, R.M., Chen, Y., Haynes, J.D., 2011. Encoding the identity and location of objects in human LOC. Neuroimage 54, 2297–2307.
Cichy, R.M., Pantazis, D., Oliva, A., 2014. Resolving human object recognition in space and time. Nat. Neurosci. 17, 455–462.
Cichy, R.M., Sterzer, P., Heinzle, J., Elliott, L.T., Ramirez, F., Haynes, J.-.D., 2013. Probing principles of large-scale object representation: category preference and location encoding. Hum. Brain Mapp. 34, 1636–1651.
Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21.
Desimone, R., Duncan, J., 1995. Selective visual attention. Annu. Rev. Neurosci. 18, 193–222.
DiCarlo, J.J., Cox, D.D., 2007. Untangling invariant object recognition. Trends Cogn. Sci. 11, 333–341.
Eurich, C.W., Schwegler, H., 1997. Coarse coding: calculation of the resolution achieved by a population of large receptive field neurons. Biol. Cybern. 76, 357–363.
Fahrenfort, J.J., Scholte, H.S., Lamme, V.A.F., 2007. Masking disrupts reentrant processing in human visual cortex. J. Cogn. Neurosci. 19, 1488–1497.
Gandhi, S.P., Heeger, D.J., Boynton, G.M., 1999. Spatial attention affects brain activity in human primary visual cortex. Proc. Natl. Acad. Sci. 96, 3314–3319.
Golomb, J.D., Kanwisher, N., 2012. Higher level visual cortex represents retinotopic, not spatiotopic, object location. Cereb. Cortex 22, 2794–2810.
Graumann, M., Ciuffi, C., Dwivedi, K., Roig, G., Cichy, R.M., 2022. The spatiotemporal neural dynamics of object location representations in the human brain. Nat. Hum. Behav. 6, 796–811.
Greene, M., Oliva, A., 2009. The briefest of glances: the time course of natural scene understanding. Psychol. Sci. 20, 464–472.
Groen, I.I.A., Dekker, T.M., Knapen, T., Silson, E.H., 2022. Visuospatial coding as ubiquitous scaffolding for human cognition. Trends Cogn. Sci. 26, 81–96.
Groen, I.I.A., Ghebreab, S., Lamme, V.A.F., Scholte, H.S., 2016. The time course of natural scene perception with reduced attention. J. Neurophysiol. 2, 931–946.
Groen, I.I.A., Jahfari, S., Seijdel, N., Ghebreab, S., Lamme, V.A.F., Scholte, H.S., 2018. Scene complexity modulates degree of feedback activity during object detection in natural scenes. PLoS Comput. Biol. 14, e1006690.
Guggenmos, M., Sterzer, P., Cichy, R.M., 2018. Multivariate pattern analysis for MEG: a comparison of dissimilarity measures. Neuroimage 173, 434–447.
Hebart, M.N., Bankson, B.B., Harel, A., Baker, C.I., Cichy, R.M., 2018. The representational dynamics of task and object processing in humans. Elife 7, e32816.
Herrero, J.L., Gieselmann, M.A., Sanayei, M., Thiele, A., 2013. Attention-induced variance and noise correlation reduction in macaque V1 is mediated by NMDA receptors. Neuron 78, 729–739.
Hillyard, S.A., Teder-Sälejärvi, W.A., Münte, T.F., 1998a. Temporal dynamics of early perceptual processing. Curr. Opin. Neurobiol. 8, 202–210.
Hillyard, S.A., Vogel, E.K., Luck, S.J., 1998b. Sensory gain control (amplification) as a mechanism of selective attention: electrophysiological and neuroimaging evidence. Philos. Trans. R. Soc. B 353, 1257–1270.
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. Nat. Neurosci. 19, 613–622.
Hopfinger, J.B., Buonocore, M.H., Mangun, G.R., 2000. The neural mechanisms of top–down attentional control. Nat. Neurosci. 3, 284–291.
Isik, L., Meyers, E.M., Leibo, J.Z., Poggio, T., 2014. The dynamics of invariant object recognition in the human visual system. J. Neurophysiol. 111, 91–102.
Itthipuripat, S., Sprague, T.C., Serences, J.T., 2019. Functional MRI and EEG index complementary attentional modulations. J. Neurosci. 39, 6162–6179.
Kaiser, D., Oosterhof, N.N., Peelen, M.V., 2016a. The neural dynamics of attentional selection in natural scenes. J. Neurosci. 36, 10522–10528.
Kaiser, D., Oosterhof, N.N., Peelen, M.V., 2016b. The neural dynamics of attentional selection in natural scenes. J. Neurosci. 36, 10522–10528.
Kaiser, D., Quek, G.L., Cichy, R.M., Peelen, M.V., 2019. Object vision in a structured world. Trends Cogn. Sci. 23, 672–685.
Kar, K., Kubilius, J., Schmidt, K., Issa, E.B., DiCarlo, J.J., 2019. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. Nat. Neurosci. 22, 974–983.
Kastner, S., Pinsk, M.A., De, Weerd P, Desimone, R., Ungerleider, L.G., 1999. Increased activity in human visual cortex during directed attention in the abscence of visual stimulation. Neuron 22, 751–761.
Kastner, S., Ungerleider, L.G., 2001. The neural basis of biased competition in human visual cortex. Neuropsychologia 39, 1263–1276.
Kay, K.N., Weiner, K.S., Grill-Spector, K., 2015. Attention reduces spatial uncertainty in human ventral temporal cortex. Curr. Biol. 25, 595–600.
Khayat, P.S., Spekreijse, H., Roelfsema, P.R., 2006. Attention lights up new object representations before the old ones fade away. J. Neurosci. 26, 138–142.
Kietzmann, T.C., Spoerer, C.J., Sörensen, L.K.A., Cichy, R.M., Hauk, O., Kriegeskorte, N., 2019. Recurrence is required to capture the representational dynamics of the human visual system. Proc. Natl. Acad. Sci. 116, 21854–21863.
King, J.R., Dehaene, S., 2014. Characterizing the dynamics of mental representations: the temporal generalization method. Trends Cogn. Sci. 18, 203–210.
Koivisto, M., Railo, H., Revonsuo, A., Vanni, S., Salminen-Vaparanta, N., 2011. Recurrent processing in V1/V2 contributes to categorization of natural scenes. J. Neurosci. 31, 2488–2492.
Kravitz, D.J., Saleem, K.S., Baker, C.I., Mishkin, M., 2011. A new neural framework for visuospatial processing. Nat. Rev. Neurosci. 12, 217–230.
Kriegeskorte, N., Diedrichsen, J., 2019. Peeling the onion of brain representations. Annu. Rev. Neurosci. 42, 407–432.
Lakatos, P., Karmos, G., Mehta, A.D., Ulbert, I., Schroeder, C.E., 2008. Entrainment of neuronal oscillations as a mechanism of attentional selection. Science 320, 110–113.
Lamme, V.A.F., Roelfsema, P.R., 2000. The distinct modes of vision offered by feedforward and recurrent processing. Trends Neurosci. 23, 571–579.
Lavie, N., 1995. Perceptual load as a necessary condition for selective attention. J. Exp. Psychol. Hum. Percept. Perform. 21, 451–468.
Lee, J., Maunsell, J.H.R., 2010. Attentional modulation of MT neurons with single or multiple stimuli in their receptive fields. J. Neurosci. 30, 3058–3066.
Luck, S.J., Woodman, G.F., Vogel, E.K., 2000. Event-related potential studies of attention. Trends Cogn. Sci. 4, 432–440.
Mangun, G.R., 1995. Neural mechanisms of visual selective attention. Psychophysiology 32, 4–18.
Martínez, A., DiRusso, F., Anllo-Vento, L., Sereno, M.I., Buxton, R.B., Hillyard, S.A., 2001. Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. Vision Res. 41, 1437–1457.
Maunsell, J.H.R., 2015. Neuronal mechanisms of visual attention. Annual review of vision science 1, 373–391.
Milner, A.D., Goodale, M.A., 2006. The Visual Brain in Action. Oxford University Press, Oxford.
Murray, S.O., Wojciulik, E., 2004. Attention increases neural selectivity in the human lateral occipital complex. Nat. Neurosci. 7, 70–74.
Noesselt, T., Hillyard, S.A., Woldorff, M.G., Schoenfeld, A., Hagner, T., Jäncke, L., Tempelmann, C., Hinrichs, H., Heinze, H.J., 2002. Delayed striate cortical activation during spatial attention. Neuron 35, 575–587.
Oliva, A., 2005. Gist of the scene. In: Neurobiology of attention. Academic press, 2005. 251-256.
Peelen, M.V., Kastner, S., 2011. A neural basis for real-world visual search in human occipitotemporal cortex. Proc. Natl. Acad. Sci. 108, 12125–12130.
Peelen, M.V., Kastner, S., 2014. Attention in the real world: toward understanding its neural basis. Trends Cogn. Sci. 18, 242–250.
Rajaei, K., Mohsenzadeh, Y., Ebrahimpour, R., Khaligh-Razavi, S.-.M., 2019. Beyond core object recognition: recurrent processes account for object recognition under occlusion. PLoS Comput. Biol. 15, e1007001.
Reddy, L., Kanwisher, N., 2007. Category selectivity in the ventral visual pathway confers robustness to clutter and diverted attention. Curr. Biol. 17, 2067–2072.
Reddy, L., Kanwisher, N.G., VanRullen, R., 2009. Attention and biased competition in multi-voxel object representations. Proc. Natl. Acad. Sci. 106, 21447–21452.
Reynolds, J.H., Chelazzi, L., 2004. Attentional modulation of visual processing. Annu. Rev. Neurosci. 27, 611–647.
Roelfsema, P.R., Lamme, V.A.F., Spekreijse, H., 1998. Object-based attention in the primary visual cortex of the macaque monkey. Nature 395, 376–381.
Seijdel, N., Loke, J., van de Klundert, R., van der Meer, M., Quispel, E., van Gaal, S., de Haan, E.H.F., Scholte, H.S., 2021. On the necessity of recurrent processing during

object recognition: it depends on the need for scene segmentation. J. Neurosci. 41, 6281–6289.
Seijdel, N., Tsakmakidis, N., De Haan, E.H.F., Bohte, S.M., Scholte, H.S., 2020. Depth in convolutional neural networks solves scene segmentation. PLoS Comput. Biol. 16, e100802..
Silver, M.A., Ress, D., Heeger, D.J., 2005. Topographic maps of visual spatial attention in human parietal cortex. J. Neurophysiol. 94, 1358–1371.
Spitzer, H., Desimone, R., Moran, J., 1988. Increased attention enhances both behavioral and neuronal performance. Science 240, 338–340.
Sprague, T.C., Serences, J.T., 2013. Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. Nat. Neurosci. 16, 1879–1887.
Squire, R.F., Noudoost, B., Schafer, R.J., Moore, T., 2013. Prefrontal contributions to visual selective attention. Annu. Rev. Neurosci. 36, 451–466.
Szczepanski, S.M., Konen, C.S., Kastner, S., 2010. Mechanisms of spatial attention control in frontal and parietal cortex. J. Neurosci. 30, 148–160.
Tang, H., Buia, C., Madhavan, R., Crone, N.E., Madsen, J.R., Anderson, W.S., Kreiman, G., 2014. Spatiotemporal dynamics underlying object completion in human ventral visual cortex. Neuron 83, 736–748.
Tang, H., Schrimpf, M., Lotter, W., Moerman, C., Paredes, A., Caro, J.O., Hardesty, W., Cox, D., Kreiman, G., 2018. Recurrent computations for visual pattern completion. Proc. Natl. Acad. Sci. 115, 8835–8840.
Tootell, R.B.H., Hadjikhani, N., Hall, E.K., Marrett, S., Vanduffel, W., Vaughan, J.T., Dale, A.M., 1998. The retinotopy of visual spatial attention. Neuron 21, 1409–1422.
Treisman, A.M., Gelade, G., 1980. A feature-integration theory of attention. Cogn. Psychol. 12, 97–136.
Ungerleider, L., Haxby, J.V., 1994. 'What' and 'where' in the human brain. Curr. Opin. Neurobiol. 4, 157–165.
van Bergen, R.S., Kriegeskorte, N., 2020. Going in circles is the way forward: the role of recurrence in visual inference. Curr. Opin. Neurobiol. 65, 176–193.
Van Voorhis, S., Hillyard, S.A., 1977. Visual evoked potentials and selective attention to points in space. Percept. Psychophys. 22, 54–62.
VanRullen, R., Thorpe, S.J., 2001. The time course of visual processing: from early perception to decision-making. J. Cogn. Neurosci. 13, 454–461.
Vaziri-Pashkam, M., Xu, Y., 2017. Goal-directed visual processing differentially impacts human ventral and dorsal visual representations. J. Neurosci. 37, 8767–8782.
Võ, M.L.H., Boettcher, S.E., Draschkow, D., 2019. Reading scenes: how scene grammar guides attention and aids perception in real-world environments. Curr. Opin. Psychol. 29, 205–210.
Wandell, B.A., Winawer, J., 2015. Computational neuroimaging and population receptive fields. Trends Cogn. Sci. 19, 349–357.
Wang, L., Mruczek, R.E.B., Arcaro, M.J., Kastner, S., 2015. Probabilistic maps of visual topography in human cortex. Cereb. Cortex 25, 3911–3931.
Wolfe, J.M., 1994. Visual search in continuous, naturalistic stimuli. Vision Res. 34, 1187–1195.
Wolfe, J.M., Võ, M.L.H., Evans, K.K., Greene, M.R., 2011. Visual search in scenes involves selective and nonselective pathways. Trends Cogn. Sci. 15, 77–84.
Wyatte, D., Jilk, D.J., O'Reilly, R.C., 2014. Early recurrent feedback facilitates visual object recognition under challenging conditions. Front. Psychol. 5, 1–10.